Desk

Vol. 43 NOVEMBER, 1952 No. 7

The Journal of Educational Psychology

UNIVERSITY OF MICHIGAN
JAN 7 1953
PERIODICAL READING ROOM

CONTENTS

Measures of the Readability of Written Communication: An Evaluation 385
GEORGE R. KLARE

The Influence of a Course on the Psychology of the Adult on Attitudes toward Old People and Older Workers 400
JACOB TUCKMAN AND IRVING LORGE

An Experimental Investigation of Theories of Reversals in Reading . 408
MORLEY KRISE

A Statistical Report on a Revision of the Miami Instructor Rating Sheet 423
A. W. BENDIG

Some Perceptual Factors Measured by Motion Picture Tests . . . 430
BENJAMIN FRUCHTER AND WILLIAM W. MAHAN

The Relationship of Literary Appreciation to Certain Measurable Factors 436
DWIGHT L. BURTON

Book Reviews 440

Published monthly except June to September
$7.00 a year in the United States, Pan American Countries and the Philippines; Canada, $7.20; other countries, $7.40. Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

STEPHEN M. COREY, *Chairman*
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

WM. CLARK TROW
University of Michigan

* Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

JACK W. DUNLAP
KARL J. HOLZINGER
HAROLD E. JONES
H. H. REMMERS
PERCIVAL M. SYMONDS
MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Stephen M. Corey, Teachers College, Columbia University, New York 27, N. Y.

Books and other materials for review, and business communications should be addressed to THE JOURNAL OF EDUCATIONAL PSYCHOLOGY, Warwick & York, Publishers, 10 E. Centre St., Baltimore 2, Md.

Manuscripts should be typed and double-spaced throughout, including quotations, footnotes, and references. In order to attain

(Continued on Inside Back Cover)

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 43 November, 1952 Number 7

MEASURES OF THE READABILITY OF WRITTEN COMMUNICATION: AN EVALUATION[1]

GEORGE R. KLARE
University of Illinois

Scientific study of the comprehensibility of written communication, through 'readability formulas,' has enjoyed a phenomenal growth in recent years. Though the importance of readability has never been doubted, and though recorded reports of attempts to study it can be traced back to 900 A. D. (*27*), little was done in the way of quantitative measurement until the early 1920's. The sudden surge of research at that time was made possible by the publication of Thorndike's first *Teacher's Word Book* (*36*).

The first work was chiefly on children's material, partly because of the nature of the *Word Book* and partly because the need for readable books could be seen most clearly in this area. With the publication in 1935 of Gray and Leary's *What Makes a Book Readable* (*19*), however, the way was opened for research in the adult area. Lorge (*26*) gave the trend toward adult measurement further emphasis by presenting a simple, practical formula that, though developed specifically for children's material, was widely used for adult material as well.

Flesch (*14*), drawing on Lorge's work, presented a simple adult readability formula in 1943 which, through its popularity, gave readability work the impetus most directly responsible for the ever-increasing interest of recent years. The publication of his revised formula (*15*) and of Dale and Chall's formula (*4*, *5*), both in 1948, then supplied two techniques highly useful for increased research.

A review of the literature on readability, however, reveals the

[1] Based on a Ph.D. dissertation by the writer, University of Minnesota, 1950. Many thanks are due Professor Donald G. Paterson for helpful suggestions and criticisms.

relatively few evaluative studies that have been made and the consequent great need for them. Unanswered questions have arisen in the areas of inter-formula comparisons, the use of outside criteria of comprehensibility, sampling, practical application, and reliability (particularly of word-list formulas). It was with these needs in mind that the following study was undertaken.

THE FORMULAS USED

Three adult formulas were selected for use in this study, the Gray-Leary (*19*), the Dale-Chall (*4, 5*), and the revised Flesch 'reading-ease' (*15*). The first of these was used because it was based on the most complete and scholarly study of readability, *What Makes a Book Readable* (*19*), that has yet appeared. The latter two were used chiefly because of their recency and widespread acceptance (*16*). Moreover, their use also makes possible a study of the recent controversy over the relative adequacy of word-list versus non-word-list formulas (*16*), since the Dale-Chall formula is of the former type and the Flesch formula of the latter, and since both formulas were based on the same criterion, the McCall-Crabbs Standard Test Lessons in Reading (*30*).[2]

Because ratings of a number of books by the children's formulas of Washburne-Morphett (*40*) and Lewerenz (*24*) were available, these formulas were also included. Though primary interest was not centered on study of the readability of children's material, inclusion of the two formulas made possible a comparison with the adult formulas already chosen.

THE READING MATERIALS RATED

Books were chosen over magazines and newspapers as the type of reading material to be rated for three reasons: greater variation in difficulty (*19*), more uniform difficulty (*23*), and greater length. It was desired, furthermore, that the books chosen for study be

---

[2] Since the time that this evaluative study was completed, Flesch published a new formula (*17*) for 'measuring the level of abstraction.' The study was not revised to include an evaluation of this formula for two reasons. In the first place, its value in predicting readability is not appreciably different from that of Flesch's revised formula (*15*), but its application is considerably more difficult (see Jenkins and Jones, *20*). In the second place, it was intended primarily as an aid in measuring the 'human interest' and not the 'reading ease' of written materials (see Flesch, *18*).

suitable for adult readers, be both fiction and non-fiction, and be of a wide range of reading difficulty.

A list of three hundred fifty books satisfactory for this study was available in *What Makes a Book Readable*, with ratings by the Gray-Leary formula provided. A sample of fifty-two books (approximately fifteen per cent) was selected for more intensive study. The sample was selected to cover the entire range of difficulty and fairly represent the list as a whole. Twenty-nine of the fifty-two books in the sample had been rated by Washburne and Morphett with their formula and included in their *The Right Book for the Right Child* (*40*); seventeen of the fifty-two had been rated by Lewerenz with his formula and included in his *Books Evaluated by Means of the Vocabulary Grade Placement Formula* (*24*). The books rated by these two formulas covered nearly the entire range of difficulty of the three hundred fifty books and were again a fair representation of them.

APPLICATION OF THE FORMULAS

In rating the sample of fifty-two books by the Flesch and Dale-Chall formulas, a sampling method was devised that would at once make possible an intensive study of sampling and meet closely the requirements of both formulas. These requirements were approximately twenty-five to thirty 100-word samples per book for the Flesch formula and a 100-word sample approximately every tenth page for the Dale-Chall formula. Taking twenty-eight systematically spaced samples from each book was found to meet these requirements and also make possible a comparison of formula ratings based on twenty-eight, fourteen, and seven samples per book.

In applying the formulas, several labor-saving devices were utilized. Dale and Chall had presented a 'multiplication table of weights' with their formula (*4*, *5*) designed to cover the product of various average sentence lengths and the regression coefficient for sentence length, and the product of various numbers of words not on the Dale list and the coefficient for non-Dale words. By adding in the regression constant, a table was developed by the writer from which formula ratings could be read directly by entering with average sentence length in words and number of non-Dale words. Farr and Jenkins (*12*) had developed such a table for use with the Flesch formula, and it was also used. Further,

since both the Flesch and Dale-Chall formulas have a sentence-length element based on complete sentences (which, of course, usually end at slightly less than or more than the one hundredth word of the samples), a table was devised for reading average sentence length. The table was entered with number of complete sentences per sample and total number of words in the sentences. Such tables as these are, of course, particularly useful in large-scale rating projects.

THE RESULTS OF THE COMPARATIVE STUDY[3]

On the basis of a comparison of the books rated in common by the five readability formulas, the approximate order of the formulas, from that which gives the highest grade level ratings to that which gives the lowest, is:

1) Washburne and Morphett;
2) Flesch reading-ease;
3) Dale and Chall;
4) Lewerenz vocabulary difficulty; and
5) Gray and Leary.

However, because grade-level ratings based on adult formulas are extrapolations from primary-grade test data, raw score ratings should preferably be used where accurate comparisons are needed; such ratings were used wherever possible in this study.

A comparison was made of the intercorrelations of the various formula ratings with each other. Tables I and II present these data.

These three intercorrelations were next checked to determine whether or not they were significantly different from each other, statistically. It was found that:

1) The Dale-Chall and Gray-Leary correlation was significantly higher than the Flesch and Gray-Leary correlation at the five per cent but not at the one per cent level;

2) The Flesch and Dale-Chall correlation was not significantly different from the Gray-Leary and Dale-Chall correlation at either the five per cent or one per cent levels; and

---

[3] A completed table of the books used in this study, with ratings by the five readability formulas, is on file at the American Documentation Institute. Order Document 3410 from the American Documentation Institute, 1719 N Street, N. W., Washington 6, D. C., remitting $1.00 for microfilm (images one inch high on standard 35 mm. motion picture film) or $1.00 for photocopies (six by eight inches) readable without optical aid.

3) The Flesch and Dale-Chall correlation was significantly higher than the Flesch and Gray-Leary correlation at both the five per cent and one per cent levels.

Again, the intercorrelations were checked for significance of differences. It was found that:

1) The Dale-Chall and Washburne-Morphett correlation was significantly higher than the Flesch and Washburne-Morphett correlation at the five per cent but not at the one per cent level;

2) The Washburne-Morphett and Gray-Leary correlation was not significantly different from zero, and therefore no further significance checks were made.

The correlations of the Flesch, Dale-Chall, Gray-Leary, and Washburne-Morphett ratings with the Lewerenz ratings were not significantly different from each other at either the five per cent or one per cent levels. (These data are not shown in table form.) The small N available (seventeen) may well have been the reason for this.

TABLE I.—INTERCORRELATIONS OF THE RATINGS OF 52 BOOKS BY THREE FORMULAS

Formula Ratings	N	r
1. Flesch and Dale-Chall	52	+.83*
2. Flesch and Gray-Leary	52	+.56
3. Dale-Chall and Gray-Leary	52	+.70

*Low scores with the Dale-Chall formula indicate easily read material, while low scores with the Flesch formula indicate difficult material. For that reason, the actual correlation coefficient is a minus value. In order to avoid confusion, this coefficient, and all others which are negative for the same reason, have been reported as positive values.

TABLE II.—INTERCORRELATIONS OF THE RATINGS OF 29 BOOKS BY FOUR FORMULAS

Formula Ratings	N	r
1. Washburne-Morphett and Flesch	29	+.40
2. Washburne-Morphett and Dale-Chall	29	+.67
3. Washburne-Morphett and Gray-Leary	29	+.19

The data from the comparative study provided other material which could be used in examining two hypotheses concerning the inaccuracies of readability formulas. Washburne and Morphett (*39*) felt that their early formula (*38*) was inaccurate in rating very easy books too hard and very difficult books too easy; this was indicated for other formulas as well by Lostutter's

(*29*) and Swanson's (*35*) studies. A second hypothesis, that sampling errors may result in inaccuracies in ratings, was suggested by the work of Dolch (*9, 10*), Elliott (*11*), and Leifeste (*22*). The tendency was found for more and/or larger samples to result in easier ratings than fewer and/or smaller samples.

In order to test the first hypothesis, the three adult formula ratings for each book were converted into standard sources. The three standard scores for each book were then compared, and the differences were found to be divided equally among three categories: 0.5 SD or less, 0.6 to 1.0 SD, and 1.1 SD or more. The occurrence of the 'large' and 'small' differences was then determined for the twenty-six 'middle-range' books and the thirteen 'easiest' plus the thirteen 'hardest,' as ranked by the Gray-Leary formula. Table III presents these data.

TABLE III.—DIFFERENCES IN THE RATINGS OF 52 BOOKS BY THE GRAY-LEARY, FLESCH, AND DALE-CHALL FORMULAS

Formula	0.5 SD or Less	1.1 SD or More
1. Gray-Leary		
26 Middle-range Books	9	6
13 Easiest plus 13 Hardest Books	9	11
2. Flesch		
26 Middle-range Books	9	8
13 Easiest plus 13 Hardest Books	9	9
3. Dale-Chall		
26 Middle-range Books	9	6
13 Easiest plus 13 Hardest Books	9	11

Though there is a tendency to find more large deviations at the extremes, in no case is the difference between large and small deviations significant at the five per cent level when the Chi-square test is used.

To test the second hypothesis, mean book ratings based on three sampling sizes were compared. Table IV presents these data. Though there is a slight tendency for the larger number of samples to provide easier ratings than the smaller number, the differences are not statistically significant.

The books rated varied in number of pages from 48 to 601; since the same number of samples (twenty-eight) was taken from each, the possibility still remained that sampling could introduce a

TABLE IV.—A COMPARISON OF THREE SIZES OF SAMPLING OF 52 BOOKS, USING THE FLESCH AND DALE-CHALL FORMULAS

Formula	Samples per Book	Mean	Standard Deviation
Flesch	28	66.4	11.8
Flesch	14	67.0	11.8
Flesch	7	67.5	12.8
Dale-Chall	28	6.79	0.92
Dale-Chall	14	6.78	0.96
Dale-Chall	7	6.76	1.02

significant error. In order to check on this, the number of 'large' deviations (i.e., from the 28-sample value for the 14-sample and 7-sample values) were computed for the thirteen longest books (365 pages or more) and the thirteen shortest books (237 pages or less). Large deviations are those which differed by one-fourth SD or more (3.0 formula points with the Flesch formula and .23 with the Dale-Chall formula). Table V presents these data. Though there is a tendency to find more large deviations

TABLE V.—NUMBER OF LARGE DEVIATIONS FROM 28-SAMPLE VALUES OF 7-SAMPLE AND 14-SAMPLE VALUES IN THE 13 LONGEST AND 13 SHORTEST OF 52 BOOKS

	13 Shortest Books	13 Longest Books
Flesch 7-Sample Values	5	7
Flesch 14-Sample Values	2	4
Dale-Chall 7-Sample Values	3	4
Dale-Chall 14-Sample Values	4	2

in longer books, the differences are not significant at the five per cent level when the Chi-square test is used.

It thus seems that differences among ratings by various formulas are probably not due to sampling errors or the ease or difficulty of the materials tested, but are rather due to such inaccuracies as inhere in the formulas themselves. A further attempt to study the comparative validity of the adult formulas was therefore made using data from adult reading tests.

CORRELATIONS WITH READING TEST DATA

The data from few reading tests are presented in a form satisfactory for use as validity criteria for readability formulas, and almost no such data are available for adult formulas. The McCall-Crabbs Standard Test Lessons in Reading, upon which the Flesch and the Dale-Chall formulas were based, are among the most satisfactory criteria available even though they were designed for use with children.

Data from two adult reading tests, however, were usable in the study. The first test is that developed by Gray and Leary for use in devising their readability formula (*19*). It is in two forms, with a total of forty-eight passages, and covers a wide range of fiction and non-fiction reading materials.

The second test is that developed by Ojemann (*31*) for use in his study of the reading ability of parents and the readability of parent-education materials. Lorge (*28*) feels that it is perhaps the only research based on truly adult materials, and Dale (*3*) believes that its subject matter (health materials) is perhaps of the most potential interest of all non-fiction materials for adults. While it should provide for a high level of interest and motivation, it is unfortunate that it consists of only sixteen passages.

Table VI presents correlations of test passage ratings, by each

TABLE VI.—CORRELATIONS OF THE RATINGS OF READING TEST PASSAGES WITH DATA BASED ON THE TESTS

	Gray-Leary Ratings	Flesch Ratings	Dale-Chall Ratings
McCall-Crabbs Test (300+ passages)	+.55*	+.71†	+.70††
Gray-Leary Test (48 passages)	+.64**	+.55	+.61
Ojemann Test (16 passages)	+.82	+.82	+.87

* Reported by Lorge (*26*)
** Reported by Gray and Leary (*19*)
† Reported by Flesch (*15*)
†† Reported by Dale and Chall (*4*)

of the three adult formulas, with readability scores for the passages as determined by objective questions based on them.

The differences between the Flesch and Dale-Chall ratings on

the Gray-Leary Test data, and the differences between the Gray-Leary, Flesch, and Dale-Chall ratings on the Ojemann Test data were not found to be significant at the five per cent level.

While Table VI shows that the Dale-Chall formula stands up slightly better than the other two formulas, the coefficients are so similar that it appears that the tests may well be measuring the same aspects of adult readability.

Table VII presents the intercorrelations of the test passage

TABLE VII.—INTERCORRELATIONS OF THE RATINGS OF THREE ADULT READABILITY FORMULAS ON THE PASSAGES OF TWO READING TESTS

Reading Test	Formula Ratings	N	r
Gray-Leary	Dale-Chall and Flesch	48	+.92
Ojemann	Dale-Chall and Flesch	16	+.98
Ojemann	Dale-Chall and Gray-Leary	16	+.83
Ojemann	Flesch and Gray-Leary	16	+.80

ratings. None of the differences in the intercorrelations of the ratings on the Ojemann test are significantly different from each other at the five per cent level.

These data indicate that the formulas are probably measuring essentially similar elements of difficulty, despite Flesch's allegation to the contrary (*16*).

FURTHER OBSERVATIONS ON READABILITY FORMULAS

In the process of rating the books and reading test passages used in the study, several pertinent observations were made regarding readability formulas in general and the Flesch and Dale-Chall formulas in particular.

The first point that requires examination is the use of 'word' factors in readability formulas. Many studies (e.g., *3*, *8*, *9*, *32*) have been extremely critical of the use of 'running counts' of words (the estimated number of different words in a piece of writing) and of the use of word-lists based on the frequency of occurrence of words. The criticisms in the literature have been sufficiently conclusive to indicate that formulas having such factors should be used very cautiously; the relatively poor showing of the Washburne-Morphett formula, though the formula was incompletely tested in this study, agrees.

The crucial question, however, is whether word familiarity factors are equally suspect. Flesch (*14*) believes so, but the data in this study indicate they are not. The previously cited studies (*3, 8, 9, 32*) showed, furthermore, that word frequency and word familiarity are by no means perfectly correlated.

Word-lists based on familiarity of words are, however, open to some rather obvious criticisms. For example, it would seem that since 'hoop' is in the Dale 3000-Word List, 'barrel' should also be, but it is not. Uncommon spellings, words misspelled for effect, dialect, etc., may well be rated too hard by their exclusion from the Dale list. Phrases may be familiar even though one or more of the words that compose them are not ('of course'); words that are unfamiliar may be counted as two familiar words if they happen to be hyphenated at the end of a printed line ('flatboat,' 'flat-boat'); words familiar in their simplest meanings may be unfamiliar in other meanings ('bear' as animal versus 'bear' as carry, or 'page' as part of a book versus 'page' as servant).

The use of a syllable count, however, does not solve this problem, for it is subject to some of the same criticisms and to certain others. Adding syllables to a word, for example, often makes it more familiar ('bout' and 'about,' or 'venture' and 'adventure'). Also, while uncommon spellings, misspellings, and dialect may be rated too hard by exclusion from a word-list, a syllable count may well rate them too easy. Some words are not easily broken into syllables without the time and effort of consulting a dictionary; others can be broken into syllables only with uncertainty (246 as 'two-forty-six' or 'two-hundred-forty-six' or 'two-hundred-and-forty-six').

Undoubtedly many other such criticisms could be found of both word-lists and syllable counts. In practice, they must be ignored. It would be possible, of course, to devise rules to handle such contingencies, but their added cumbersomeness could not be justified by the small increase in validity they would make possible.

The practical matter of the time necessary for application of the formulas is another on which the comparative study provided data. (The data are tentative because they are based on the work of only one person—the writer—and on a number of books of varying reading difficulty and type legibility.) Such data, though tentative, have long been presented. Lively and

Pressey (*25*) estimated that it took three hours to rate a book with their formula; Johnson (*21*) estimated that it took but thirty minutes to rate a book with his formula; and Fihe, Wallace, and Schulz (*13*) estimated that it took an average of six minutes to rate a 100-word sample with the Flesch formula (approximately two and one-half to three hours per book).

In the present study, several conclusions were reached. In the first place, it is difficult to determine a definite time necessary for rating a book, not only because of variations in books themselves but also because the rater's skill increases with practice. In the second place, the use of prepared tables (such as those of Farr and Jenkins (*12*) or the writer's) greatly diminishes the time necessary for rating, particularly if a number of books are rated.

In the third place, Flesch formula ratings require a slightly smaller amount of time than the Dale-Chall ratings on initial usage, but about the same time after several books are rated. The average time found by the writer when using prepared tables was about one and three-quarters to two hours per book (about four minutes per 100-word sample) for both formulas. It was further found that after considerable usage (rating about fifteen books) and when making a special attempt at speed it took slightly less time to rate a book with the Dale-Chall than with the Flesch formula.

THE FUTURE

With the rapidly growing acceptance of readability measures has come a feeling on the part of some critics that formulas are often overextended. One can easily see from the consideration of satire, irony, subtle humor, etc., that the problem of readability is a more complex one than the simple methods of today would lead one to believe. Readability formulas do not predict reading difficulty perfectly, and they must never be used with the tacit assumption that they do.

Future readability research aimed at increasing prediction can take two directions. The first is increased complexity. Dale and Chall (*6*), for example, believe that formulas can and will at some future time include factors of format and organization, content and expressional elements. The problem should be investigated, but only experience will indicate whether or not the probably small increase in validity provided will justify

increased formula complexity and the consequent greater time and effort necessary to application (see *20*).

The second direction is to increase the accuracy of the shorthand methods of today. The greatest need here is for an accurate criterion against which to validate formulas, that is, an improved adult reading test. The time and expense necessary in developing such a test makes its appearance in the near future unlikely.

Much can be done with available research, however, in the way of utilizing the data gathered by the many other disciplines primarily concerned with language and communication, such as linguistics, semantics, typography, the psychology of reading and of attitudes and opinions, etc. Examples of recent research not primarily concerned with readability indicate the value of this approach. The first is a reading study by Dearborn, Johnston, and Carmichael (*7*) which shows that the ability of a subject to understand a passage can be given in quantitative terms by noting the words he stresses when reading aloud. The authors suggest that an arrangement which indicates emphasis (bold-face type, italics, etc.) should be used more freely in written communication. Trager (*37*) and Calkins (*2*) believe that, while emphasis is important, the arrangements suggested, or other such artificial arrangements, would eventually defeat their purpose; the former suggests the further study of linguistic structure and analysis, and the latter more competent use of punctuation and syntax as better alternatives.

A somewhat more radical departure is suggested by the work of Andrews (*1*). He found that a new kind of topography he calls 'square span' increases reading speed and comprehension. An example of it is this sentence:

Notice	how	these	are
	easily	units	visualized.

These examples are only indications of the fruitfulness of an interdisciplinary approach to readability problems. It is not unlikely that, as the science of readability progresses, it will one day relate itself to such work as that of Shannon (*34*) on the technical (engineering) aspects of communication. Today's progress in improving the centuries-old methods of written communication is indeed gratifying; the progress of tomorrow should be even more so.

SUMMARY AND CONCLUSIONS

An evaluative study was made of several formulas for measuring the readability of written materials. It was based on an analysis of fifty-two books and three reading tests. These conclusions arose from the study:

1) Correlations of book ratings by the Gray-Leary, Flesch, and Dale-Chall formulas show that the three measure much the same aspects of adult readability, and that the Dale-Chall and Flesch formulas are most similar; there is some slight evidence that the Dale-Chall formula is superior to the other two.

2) Correlations of formula ratings with reading test data also show that the three formulas measure much the same aspects of adult readability and that the Dale-Chall and Flesch formulas are most similar; the differences between the three are, however, very slight.

3) The differences in formula ratings do not appear to be due to ease or difficulty of the materials measured or to errors involved in sampling, but rather to be due to the inadequacies of, and differences between, the reading tests on which they were based and to failure of the formulas to measure all of the significant aspects of readability.

4) Readability formulas are, however, sufficiently accurate for estimating the comparative readability of adult materials.

5) Formula ratings in terms of estimated reading grade levels, such as those presented by the Washburne-Morphett and Lewerenz children's formulas, are necessary for children's materials for practical reasons; such grade ratings by adult formulas are of questionable meaning for adult materials.

6) Recent adult readability formulas are relatively simple and take little time to apply. The use of prepared tables and of smaller book samples greatly diminish the time and effort necessary for formula application without materially affecting the accuracy of ratings.

7) Perhaps the greatest need for future improvement of readability formulas is an accurate adult reading test; a thorough interdisciplinary approach using available data gathered by the various sciences concerned with communication should suggest a number of ways of improving present methods of measuring readability.

BIBLIOGRAPHY

1) R. B. Andrews. "Reading power unlimited." *Texas Outlook*, 1949, 33, 20–21.

2) F. C. Calkins. "Comment." *Science*, 1950, 111, 582–583.

3) E. Dale. "Evaluating Thorndike's word list." *Educ. Res. Bull.*, Ohio State Univ., 1931, 10, 451–457.

4) E. Dale and Jeanne S. Chall. "A formula for predicting readability." *Educ. Res. Bull.*, Ohio State Univ., 1948, 27, 11–20.

5) ———. "A formula for predicting readability: instructions." *Educ. Res. Bull.*, Ohio State Univ., 1948, 27, 37–54.

6) ———. The concept of readability. *Elem. Engl.*, 1949, 26, 19–26.

7) W. F. Dearborn, P. W. Johnston and L. Carmichael. "Oral stress and meaning in printed material." *Science*, 1949, 110, 404.

8) E. W. Dolch. *Reading and word meanings.* Boston: Ginn and Co., 1927.

9) ———. "Vocabulary burden." *J. educ. Res.*, 1928, 17, 170–183.

10) ———. "Sampling of reading matter." *J. educ. Res.*, 1930, 22, 213–215.

11) Catherine J. Elliott. "A critical analysis of the objective method of measuring reading difficulty." *Pittsburgh Schools*, 1941, 15, 201–209.

12) J. N. Farr and J. J. Jenkins. "Tables for use with the Flesch readability formulas." *J. appl. Psychol.*, 1949, 33, 275–278.

13) Pauline J. Fihe, V. Wallace and N. Schulz (Ed.). *Books for adult beginners, grades I to VII* (Rev. Ed.). Chicago: American Library Ass'n., 1946.

14) R. F. Flesch. *Marks of readable style.* Columbia Univ. Contributions to Education, No. 897. New York: Bureau of Publications, Teachers College, Columbia University, 1943.

15) ———. "A new readability yardstick." *J. appl. Psychol.*, 1948, 32, 221–233.

16) ———. "A dissenting opinion on readability." *Elem. Engl.*, 1949, 26, 332–334, 340.

17) ———. "Measuring the level of abstraction." *J. appl. Psychol.*, 1950, 34, 384–390.

18) ———. "Reply to criticism by Jenkins and Jones." *J. appl. Psychol.*, 1951, 35, 69.

19) W. S. Gray and Bernice Leary. *What makes a book readable.* Chicago: The Univ. of Chicago Press, 1935.

20) J. J. Jenkins and R. L. Jones. "Flesch's 'Measuring the Level of Abstraction.'" *J. appl. Psychol.*, 35, 68.

21) G. R. Johnson. "An objective method of determining reading difficulty." *J. educ. Res.*, 1930, 21, 283–287.

22) Bertha V. Leifeste. "An investigation of the sampling of reading materials." *J. educ. Res.*, 1944, 37, 441–450.

23) A. S. Lewerenz. "Vocabulary grade placement of typical newspaper content." *Los Angeles Educ. Res. Bull.*, 1930, 10, 4–6.

24) ———. *Books evaluated by means of the vocabulary grade placement formula* (Revised to 1937). Los Angeles, California: Los Angeles City School District, 1937.

25) Bertha A. Lively and S. L. Pressey. "A method for measuring the 'vocabulary burden' of textbooks." *Educ. Admin. and Supv.*, 1923, 9, 389–398.

26) I. Lorge. "Predicting reading difficulty of selections for children." *Elem. Engl. Rev.*, 1939, 16, 229–233.

27) ———. "Word lists as background for communication." *Teach. Coll. Rec.*, 1944, 45, 543–552.

28) ———. "Readability formulae—an evaluation." *Elem. Engl.*, 1949, 26, 86–95.

29) M. Lostutter. "Some critical factors of newspaper readability." *Journalism Quart.*, 1947, 24, 307–314, 331.

30) W. A. McCall and Lelah Mae Crabbs. *Standard test lessons in reading.* New York: Bureau of Publications, Teachers College, Columbia University, 1926.

31) R. H. Ojemann. *The reading ability of parents and factors associated with reading difficulty of parent education materials.* Univ. of Iowa Studies in Child Welfare, 1934, 8 (I), 11–32.

32) H. Richardson. *A study of the vocabulary used in lecturing to a general college class.* Unpublished ms., Univ. of Minn., 1947.

33) W. Schramm. (Ed.). *Communications in modern society.* Urbana, Illinois: Univ. of Illinois Press, 1948.

34) C. E. Shannon. "A mathematical theory of communication." *Bell Syst. Tech. J.*, 1948, 27, 379–423 and 623–656.

35) C. Swanson. "Readability and readership: a controlled experiment." *Journalism Quart.*, 1948, 25, 339–343.

36) E. L. Thorndike. *The teacher's word book.* New York: Bureau of Publications, Teachers College, Columbia University, 1921.

37) G. L. Trager. "Prosodic patterns and oral stress." *Science*, 1950, 111, 582.

38) Mabel Vogel and C. Washburne. "An objective method of determining grade placement of children's reading material." *Elem. Sch. J.*, 1928, 28, 373–381.

39) C. Washburne and Mabel Vogel Morphett. "Grade placement of children's books." *Elem. Sch. J.*, 1938, 38, 355–364.

40) ———. *The right book for the right child* (Third Ed.). New York: The John Day Co., 1942.

THE INFLUENCE OF A COURSE ON THE PSYCHOLOGY OF THE ADULT ON ATTITUDES TOWARD OLD PEOPLE AND OLDER WORKERS*

JACOB TUCKMAN and IRVING LORGE

Teachers College, Columbia University

In previous studies two questionnaires were utilized to investigate the attitudes of one hundred forty-seven graduate students, twenty to fifty-one years of age, toward old people and toward the older worker (*1*, *2*). The first questionnaire consisted of one hundred thirty-seven statements about old people classified into thirteen categories: Physical, Financial, Conservatism, Family, Attitude toward Future, Insecurity, Mental Deterioration, Activities and Interests, Personality Traits, Best Time of Life, Sex, Cleanliness, and Interference. The second questionnaire consisted of fifty-one statements about older workers classified into nine categories: Resistance to New, Physical, Reaction to Criticism, Mental, Keeping Youth Down, Employer Attitudes—Costs, Waiting for Retirement, Interpersonal, and Jobs. With the exception of statements dealing with physical changes with age, most of the statements in the two questionnaires were beliefs and generalizations about old people and the older worker for which there is no valid basis. The findings showed that even a sophisticated group of graduate students agree substantially with these beliefs and generalizations about old people and the older worker, indicating that old age is looked upon as a period of economic insecurity, poor health, loneliness, and failing physical and mental powers. These findings would suggest that old people and older workers are living in a social climate which is not conducive to good adjustment in their later years and points up the need for more objective data to prove or disprove the prejudices and misconceptions about the abilities and physical and personality characteristics of old people and the older worker.

---

* *Retirement and Adjustment Series:* Number 7. Sponsored coöperatively by The Institute of Adult Education and The Institute of Psychological Research, Teachers College, Columbia University.

The purpose of this study is to determine the influence of a course on the psychology of the adult in modifying the attitudes of graduate students toward old people and older workers.

At the first session of the course the two questionnaires described above were administered to the graduate students referred to in the previous studies. At the completion of the course, a shorter questionnaire of forty statements was administered to the same group. This questionnaire consisted of thirty statements selected from the original one hundred thirty-seven statements about old people and ten from the original fifty-one statements about the older worker. All statements showing age or sex differences on the initial test were selected. There were twenty such statements—seven showing age differences and thirteen showing sex differences. The remaining twenty questions were selected at random from different sections of the original questionnaires. The shorter questionnaire was administered to the group as part of the final examination for the course. The directions were the same except for the following additional statement: "Your replies to this questionnaire will have no bearing on your final mark for this course."

Test records for initial and final tests were available for eighty men and forty-four women. Age comparisons of the responses to the forty statements at the beginning and end of the course were made for students under thirty years of age and those over thirty years of age. The group under thirty years of age included thirty-nine men and eighteen women. The mean age of the group was 24.6 years with a standard deviation of 2.2 years. The group over thirty years of age included forty-one men and twenty-six women. The mean age of this group was 36.3 years with a standard deviation of 5.1 years.

The mean number of 'yes' responses for the forty statements on the initial and final tests by sex and age is given in Table 1. From the data in Table 1 it is evident that there is a tendency for both age groups to agree more with the beliefs and generalizations about old people and the older worker on the final than on the initial test. However, the difference between initial and final tests is not significant for either sex under thirty years of age, but is significant at the one per cent level for both men and women over thirty years of age. An analysis of the difference between the mean number of 'yes' responses on initial and final tests,

employing a finer age breakdown, illustrates more clearly the tendency for the agreement with the beliefs about old people and the older worker to increase with age. The test-retest difference in the mean number of 'yes' responses is 1.11 for twenty to twenty-four year olds, .03 for twenty-five to twenty-nine year olds, 2.74 for thirty to thirty-four year olds, 4.40 for thirty-five to thirty-nine year olds, and 5.69 for forty year olds and over. The correlation between age and difference between initial and final tests is .24 for men, .26 for women and .26 for the total group. The correlation is not significantly different from zero for women but is significant at the five per cent level for men and at the one per cent level for the total group.

TABLE 1.—MEAN NUMBER OF 'YES' RESPONSES AND STANDARD DEVIATION FOR TEST AND RETEST ON OLD PEOPLE AND OLDER WORKER QUESTIONNAIRES, BY AGE

Group	N	Test		Retest		Test-Retest	
		Mean	SD	Mean	SD	t	r
Male 20–29	39	16.4	5.5	16.8	6.5	.4	.47
Male 30–51	41	15.4	5.8	19.1	6.6	3.3*	.36
Female 20–29	18	17.8	6.8	18.7	7.9	.5	.62
Female 30–51	26	17.5	8.2	22.0	7.0	3.4*	.62
All 20–29	57	16.9	6.0	17.4	7.0	.7	.54
All 30–51	67	16.2	6.9	20.2	6.9	4.7*	.50

* Significant at .01 level.

The test-retest correlations range from .36 to .62. Employing the Spearman-Brown prophecy formula the reliability of the one hundred eighty-eight statements in the two original questionnaires ranges from .73 for men over thirty to .88 for women over thirty. For the total group the reliability is .82. These reliability coefficients may be unduly high because the method of selection for the shorter questionnaire may have favored the inclusion of the more reliable items.

An analysis of the forty statements was made for initial and final tests in order to determine any changes in response to

TABLE 2.—PROPORTION OF TEACHERS COLLEGE GRADUATE STUDENTS, BY AGE, IN AGREEMENT WITH STATEMENTS ABOUT OLD PEOPLE AND OLDER WORKERS

Category	Statement	Age Group				Significant Difference Between:
		1 20–29 N = 57 Test %	2 30–51 N = 67 Test %	3 20–29 N = 57 Retest %	4 30–51 N = 67 Retest %	
Old People						
Physical	4. They need less sleep than younger people.	30	54	33	48	1–2
	8. They feel cold even in warm weather.	40	46	67	67	1–3, 2–4
	12. They die soon after retirement.	33	46	35	55	3–4
	17. They walk slowly.	86	87	88	94	
	26. They develop infection easily.	60	37	72	70	1–2, 2–4
	30. They die after a major operation.	25	10	39	25	2–4
	Mean Physical	46	47	56	60	
Mental Deterioration	1. They are absent-minded	26	43	36	39	1–2
	18. They cannot remember names.	46	40	35	55	2–4, 3–4
	22. They like to doze in a rocking chair.	51	57	53	66	
	23. They are in their second childhood.	14	15	18	16	
	29. They frequently talk to themselves.	42	37	31	45	
	Mean Mental Deterioration	36	38	35	44	
Family	3. They are proud of their children.	93	91	93	93	
	5. They are not important in family affairs.	12	18	14	25	
	7. They are a burden to their children.	37	22	47	42	2–4
	28. They get love and affection from their children.	72	69	63	70	
	Mean Family	54	50	54	58	

Activities and Interests	6. They are more interested in religion.	81	73	91	90	2–4
	16. They like religious programs on the radio.	61	52	65	69	2–4
	19. They collect many useless things like string, paper and old shoes.	28	34	21	34	
	20. They like to play checkers or dominoes.	49	42	47	58	
	Mean Activities and Interests	55	50	56	63	
Conservatism	10. They are conservative.	89	88	84	87	
	11. They are out of step with the times.	42	39	33	51	3–4
	21. They object to women smoking in public.	61	40	47	52	1–2
	24. They prefer old friends rather than make new ones.	80	76	82	88	
	Mean Conservatism	68	61	62	70	
Best Time of Life	2. They are in the happiest period of their lives.	23	6	7	1	1–2, 1–3
	15. They never had it better.	7	15	5	3	2–4
	Mean Best Time of Life	15	11	6	2	
Financial	13. They have too much power in business and politics.	35	25	33	30	
	27. They are usually supported by their children or old age pensions.	72	58	70	75	2–4
	Mean Financial	54	42	52	53	
Attitude Toward Future	9. They think the world is headed for destruction.	11	25	19	40	1–2, 3–4
	Mean Attitude toward the Future	11	25	19	40	
Personality Traits	14. They like to give advice.	82	85	90	88	
	Mean Personality Traits	82	85	90	88	
Insecurity	25. They frequently are at loose ends.	31	51	54	49	1–2, 1–3
	Mean Insecurity	31	51	54	49	
Older Worker						
Physical	31. Older workers are slow.	63	67	54	78	3–4

TABLE 2.—(*Continued*)

Category	Statement	Age Group				Significant Difference Between:
		1 20–29 N = 57 Test %	2 30–51 N = 67 Test %	3 20–29 N = 57 Retest %	4 30–51 N = 67 Retest %	
Older Worker (Cont.)						
Physical	37. They get occupational diseases more often.	28	21	39	45	2–4
	39. They are not physically able to keep up with the work.	19	18	19	24	
	Mean Physical	37	35	37	49	
Mental	33. They are in a rut.	35	22	19	39	1–3, 2–4, 3–4,
	38. They cannot concentrate.	9	6	9	19	2–4
	Mean Mental	22	14	14	29	
Interpersonal	34. They cannot supervise others well.	11	9	9	13	
	40. They are difficult to work with.	9	12	16	21	
	Mean Interpersonal	10	11	13	17	
Resistance to New	32. They fail to keep up with changing methods of work.	39	33	37	52	2–4
	Mean Resistance to New	39	33	37	52	
Employer Attitudes —Costs	36. They increase costs of pensions for employers.	46	49	51	63	
	Mean Employer Attitudes—Costs	46	49	51	63	
Waiting for Retirement	35. They will not take on additional responsibilities	16	7	19	33	2–4
	Mean Waiting for Retirement	16	7	19	33	

individual statements, to identify those statements which may have contributed to the gain in the mean number of 'yes' responses, and to determine any differences in the performance of the two age groups. The forty statements about old people and the older worker under the categories into which they had been classified, showing the proportion of 'yes' responses on initial and final tests for the two age groups are presented in Table 2.

The data in Table 2 show a shifting of age differences. There were seven statements which showed significant age differences on the initial test. All but one of these differences disappear on the retest. However, on the final test there were five statements showing significant age differences which were not evident on the initial test.

In comparing the proportion of 'yes' responses for initial and final tests, four statements in the questionnaires show a significant difference for the group under thirty years of age, and fourteen for the group over thirty years of age. For the group under thirty all but one of these items show an increase in negative attitude toward old people and the older worker; for the group over thirty years of age all show this tendency.

DISCUSSION

It is evident that a course on the psychology of the adult has little effect on the attitudes of graduate students toward old people and the older worker. Certain shifts do occur such as the leveling off of age differences in responses to some statements and the emergence of age differences in responses to others. For the majority of statements there is no change and the data, if anything, indicate that the beliefs and stereotypes about old people and the older worker are reinforced. This reinforcement may be attributed almost entirely to the group over thirty years of age. The increase in stereotypy may lie in the orientation of the course itself and in the motivation of the graduate students who enrolled in it. The course was not geared specifically to dispel the beliefs and stereotypes about old people and the older worker, but rather to give objective information on the physical changes with age, such as decline in visual and auditory acuity, and strength, and on the psychological changes such as reduction in range of adjustment. The course emphasized that intellectual power does not decline, and that it was important for the older

person to obtain satisfaction out of his experiences, which implies being respected and having the approval of his friends, family and employer. A number of students in the group over thirty may have been motivated to enroll in a course on the aging process because of some concern with their own aging or that of an aging parent or relative, although it is recognized that the selection of any course of study depends on a number of other factors such as the convenience of the hour, the instructor, etc. In this connection it is interesting to note that sixteen per cent of the male students and thirty-five per cent of the female students between the ages of thirty and thirty-nine years consider themselves to be middle-aged rather than young. Since these students had accepted the cultural stereotypes about age even before instruction, the objective course material on physiological and psychological change may have only served to reinforce the concerns or fears that they may have had about their own aging, and offset any of the positive aspects of the material presented with the result that they view the future with misgiving. Since the individual's reaction to the facts about the aging process may be deleterious rather than constructive, some consideration needs to be given to developing a course in this area which will help the adult understand his own aging as a natural process and to improve his outlook on the future.

REFERENCES

1) J. Tuckman and I. Lorge. "Attitudes Toward Old People." *Journal of Social Psychology.* (in press)

2) ———. "Attitudes Toward Older Workers." *Journal of Applied Psychology*, 1952, 36, 149–153.

AN EXPERIMENTAL INVESTIGATION OF THEORIES OF REVERSALS IN READING

MORLEY KRISE

Remedial Reading Clinic of Seattle

The tendency to reverse letters and words in reading, for example, reading *b* for *d*, *d* for *q*, *p* for *b*, *q* for *p*, *on* for *no*, and *was* for *saw*, seen especially in children's reading, has been attributed to a number of causes. The study reported here is an experimental investigation of some of these theories.

The problem of determining the cause of reversals in reading has been greatly oversimplified by the common belief that all reversed letters and words are the result of rotation on their vertical axes, that they are all mirror-images of the correct letters and words. This error was recognized by Wechsler and Pignatelli (*22*), who indicate that not all of the reorientations are of this type, nor as simple as this. They point out that: " . . . when *b* is read as *d* the visual phenomenon involved is quite different from that taking place when *n* is read as *u*, and considerably less complicated than what occurs when *N* is read as *Z*. The process common to all three is that the confused letter in each case has been rotated about some imaginary axis; *b* to be changed to *d* has been rotated about its vertical axis; *n* when changed to *u* has been rotated about its horizontal axis [Actually this is an error; *n* must be rotated on its depth axis to be read as *u*. But another example could have been chosen, *p* and *b*, for *p* when changed to *b* has been rotated about its horizontal axis.]; and *N* to produce *Z* has been rotated about its depth axis. Furthermore, while *b* and *p* required a rotation of 180 degrees, *N* needed only a rotation of 90 degrees; in fact, had it been rotated 180 degrees it would have resumed its original form, or become *N* again. It thus appears that in the analysis of reversals one has not only to deal with the fact of rotation, but one must also consider the plane in which this rotation takes place, and at times, also, the angular distance involved . . . "

Cole (*4*) contends that " . . . the reversal of words or letters . . . is not an indication of anything except immaturity." Betts (*1*) shows somewhat the same point of view. Gates (*8*) sees, among other causes, the transfer to reading of the techniques

or habits of studying or recognizing other objects. Gates and Bennett (*10*) expound the same theory in an earlier writing. Wolfe (*24*) found evidence to the contrary when he conducted an experiment to test the theory that visual impressions leading to reversal responses are formed in the right-left direction.

Although a number of workers, including Hildreth (*13*) and Woody and Phillips (*25*), have investigated the possibility of an association between left-handedness *per se* and reversals and have found negative evidence, a search of the literature did not disclose any writer who held to this association. Orton (*19*) originated a theory that reversals are due to failure to have established hemispherical dominance, but it appears from the literature that he has had no 'takers' since he first introduced the theory in 1928. Another theory which has had rather broad circulation but has not been supported by the experimental literature is that reversals in reading are caused by mixed cerebral dominance, which manifests itself in opposite handedness and eyedness, that is, right-handedness and left-eyedness or left-handedness and right-eyedness. Witty and Kopel (*23*), Hildreth (*12*), Gates and Bond (*11*), and Monroe (*18*) found evidence that contradicts this theory. House (*14*) holds to it but offers no evidence in support.

Krise (*15*) suggests that reversals in reading may be due to a normal difficulty in space perception, a lack of familiarity with the proper relationship between these reversed symbols and their ground. He demonstrates this difficulty by pointing out that when the symbol b is written, the most expert reader does not know whether it is a "p" or a "d," because he does not know the intended relationship between this symbol and its ground. Fabian (*6*) disclosed a marked tendency in young children to verticalize—orient to a vertical position—horizontally oriented figures. He believes that this explains reversals in reading, but he does not offer any explanation as to how it explains them.

Fernald (*7*) does not hold to any of the above theories, saying, "There is no particular psychological problem involved in the failure to distinguish between two things that resemble each other closely. . . . " McCullough, Strang, and Traxler (*17*) also hold to none of these explanations, observing that "Various causes have been attributed to these tendencies. . . . For every exponent of a theory, however, there seems to be an investigator who knows of exceptions to the theory."

Gates (*9*), in addition to his theory of transfer of direction of inspection to reading, suggests that there are certain constitutional factors which may predispose the pupil toward the formation of inappropriate techniques—confusions, reversal, and other errors—, various visual defects which result in vague images, eye-muscle imbalance, and alternating vision being "quite certainly genuine factors."

Here we find reversals in reading attributed in turn to lack of maturation, habitual direction of inspection, left-handedness, failure to establish hemispheral dominance, mixed dextrality, normal difficulties in space perception, verticalization tendencies, lack of familiarity with the symbols involved (Fernald (*7*)), and visual defects. The experiments to be reported here were set up to investigate the degree of association, if any, between the tendency to commit reversals in reading and each of these factors, except failure to establish hemispheral dominance and verticalization tendencies, the former being excepted because it is based on neurological functions the presence or absence of which cannot be established, the latter being excepted because no technique was available for appropriately measuring this tendency.

EXPERIMENTAL INVESTIGATION

Subjects Used.—The subjects for this study were twenty adults, seven men and thirteen women, ranging in age from twenty-nine to fifty-two years. One had left school after completing the eighth grade, two on completing the tenth, seven after being graduated from high school, two after cor pleting two years of college, five after four years of college, and three after completing one year of postgraduate work in college. The reading level of all subjects except one was in the college range, this one showing an eleventh-grade level, as determined by the Diagnostic Examination of Silent Reading Abilities, Senior Division (*20, 21*). All were willing subjects and coöperated in every way possible.

Measurement of Reversal Tendencies.—This study proceeded on an assumption, later established as a fact, that all subjects when faced with the requirement of reading unfamiliar, reversible symbols would commit reversals in this reading and continue to do so with diminishing frequency until no more reversals were made.

Only single-letter reversals were used here: whole-word reversals were not investigated. In order to understand the

relationship between the symbols used and letters reversed in reading it will be necessary to arrive at a precise definition of reversals in reading. In the past much has been read into this phenomenon. Actually, when a reversal is made in reading, all we know from the standpoint of the reader is that he has miscalled a symbol or a group of symbols, and all we know from the standpoint of the symbols is that when they are reoriented or rearranged they become the symbols called. Therefore, we arrive at the following definition: When, in oral or silent reading, a letter or word or group of words is miscalled and a reorientation of the letter or a reordering of the letters within the word or the words within the group would produce the letter, word, or words called, or a close resemblance to them, this error is called a reversal in reading.

The symbols used here, with alphabetical meaning assigned to each, are shown in Fig. 1. These symbols were so designed as to

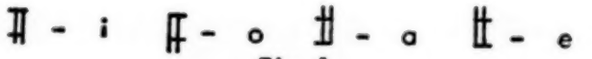

Fig. 1

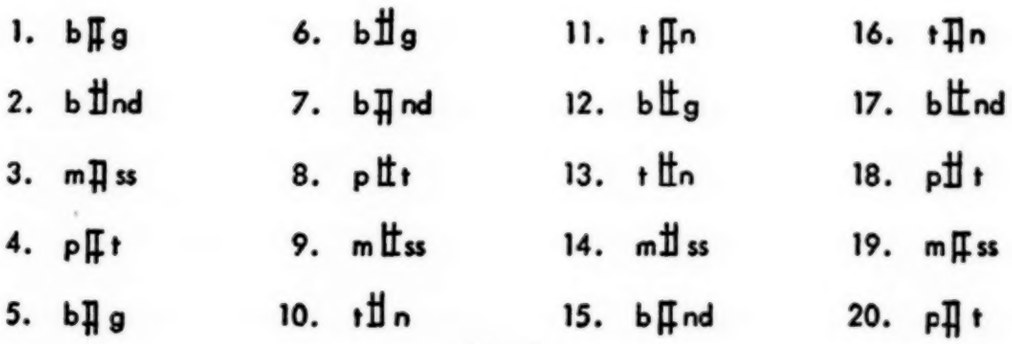

Fig. 2

be relatively unfamiliar, reversible, that is, a particular reorientation of any one will produce one of the others, and to duplicate the reversal characteristics of the letters *b*, *d*, *p*, and *q*, that is, when one of them is rotated 180 degrees on one of its three axes one of the other three symbols is produced.

The subjects were permitted to study these symbols with their equivalent letters in successive ten-second sessions, having been

told beforehand that they would be required to recognize these symbols in words in which they had been substituted for their equivalent letters. At the end of each study session they were each given a copy of the words shown in Fig. 2 and instructed to read them silently and indicate in writing their reading of each word.

The words were presented in a different order each time. Words were so selected for this list as to duplicate the situation present in the common letter reversals where the miscalling of a letter in a word still produces a familiar word, as seen, for example, in dot and pot. In the list, for example, misreading of the symbol for *o* in bog would produce bag or beg or big.

This procedure was continued by each subject in the group until he had written correctly all words in a list. Of the total number of words written by any one subject the proportion in which he misread a symbol as one of the other three was considered his index of reversal tendency.

Tests Administered.—Each subject was examined for reversal tendencies in reading familiar symbols by having him read thirty isolated reversible words and an eight-line paragraph in which these thirty words appeared in context (*10*).[1]

Maturation was determined by the chronological age of each subject; it being considered that a person over the age of twenty-five was physiologically and neurologically mature.

Habitual direction of inspection was determined in the following manner: Ten familiar objects were placed on a mantle behind a screen. The screen was then removed and the subjects were asked to list the names of the items, having been told beforehand that it was a test of identification to see how they identified familiar objects. In order not to suggest a right-to-left direction of naming the objects, the subjects were given answer sheets with numbers appearing in a column. This procedure was followed three times with a different set of objects used each time.

The degree of left-handedness was measured by the Habit Scale For Handedness (*16*), in which fifty acts which require preference for one hand or the other are presented and the subject is asked to indicate on a scale of ten points the degree to which he habitually

---

[1] All of these words were taken from the reference cited, omitting the three duplicates appearing there.

prefers the left or right hand in performing these acts. The score is computed on a percentage basis, so that 100 minus the score for the left hand gives the score for the right hand. The degree of right-handedness was used in this study for statistical convenience.

In addition to securing a measurement of the degree of left- or right-handedness, each subject was designated as discreetly left- or right-handed on the basis of his score, in order to secure data that would be comparable to those used in other studies of reversals.

Mixed eye-hand dominance is a combination of opposite handedness and eyedness. Handedness had been disclosed by the Habit Scale for Handedness. Eyedness was determined by the subjects being given the following instructions: "Place the eraser of your pencil on the tip of your nose. With both eyes open point the lead at some object across the room. While keeping the pencil in this position look at the lead with the left eye only, right eye closed. Now look at it with the right eye, left eye closed. Do this several times, removing and replacing the pencil and pointing it at different objects. Always point it with both eyes open, then look at it with each eye. If while looking at the lead with the right eye it still points at the object, you are right-eyed. If while looking at it with the left eye only it still points at the object, you are left-eyed. If neither eye alone sees the lead as pointing at the object, you are mixed-eyed. Indicate which you are in the space provided on your answer sheet."

The vision of each subject was examined by use of the Keystone Opthalmic Telebinocular and Betts' Visual Sensation and Perception Tests (*1*). In addition to supplying information on total visual efficiency for reading, Betts' Visual Sensation and Perception Tests offer data in each of the following areas: two-eyed vision, distance fusion, two-eyed efficiency, left-eyed efficiency, right-eyed efficiency, vertical balance, coördination level, distance lateral balance, reading lateral balance, reading fusion, reading sharpness of image, and distance sharpness of image. These data were secured on each subject.

Each subject's perception of space relations was determined by administering the Survey of Space Relations Ability, Form A (*2*, *3*). As standard scores on this test are given in percentile ranks the raw scores were used in statistical computations in this study.

As stated above, the reading level of each subject was determined by administering the Diagnostic Examination of Silent Reading Abilities, Senior Division (*20, 21*).

RESULTS OBTAINED

None of the subjects committed reversals while reading the familiar reversible symbols in the initial test of tendency to reversals. All subjects made reversals in their reading of the relatively unfamiliar reversible symbols designed for this study, ranging from three per cent to seventy-one per cent of the test words attempted.

As stated earlier, all subjects were considered to be physiologically and neurologically mature, as the youngest subject was twenty-nine years of age.

Correlation coefficients were computed to determine the relationship between the indexes of reversal tendencies obtained from the subjects' reading of test items and all other factors for which tests were administered. All measurements made by Betts' tests are reported in a dichotomy—the presence or absence of the particular visual defect. Although reported in a dichotomy, these defects can be considered to be continuous and normally distributed. Therefore, biserial coefficients of correlation were computed for them. This was the case also with handedness, eyedness, and eye-hand dominance. Product-moment correlation coefficients were computed for all other factors.

As the scores on all of the tests except per cent of left-handedness are such that a high score is considered 'good,' the indexes of reversals were reported on the same basis, 100 per cent indicating a complete absence of reversal tendency. For the same reason, per cent of right-handedness was recorded, rather than its reciprocal, per cent of left-handedness.

All of the subjects listed the objects in the Direction of Inspection Tests in their left-to-right order as they appeared on the mantle. It is believed that this was done out of habit rather than insight into the true purpose of the test, as all subjects were precise, formal, and highly technical in some of their responses, suggesting that they believed this to be a test of identification, as they were told in the instructions.

In order to be considered statistically significant at the five and at the one per cent level a coefficient of correlation based on

the number of measures secured here—twenty pairs—would need to be of the magnitude of .444 and .561 or greater, respectively. In other words, if, when based on twenty pairs of measures, a correlation coefficient is .444 or greater, the chances are ninety-five per cent or greater that this correlation coefficient did not occur by pure chance, that it represents a true relationship. In like manner, if the coefficient is .561 or greater, the chances are ninety-nine per cent or greater that it represents a true relationship (*5*).

TABLE I.—RESULTS OF TESTS OF REVERSAL TENDENCY, DIRECTION OF INSPECTION, HANDEDNESS, EYEDNESS, EYE-HAND DOMINANCE, AND SPACE PERCEPTION, WITH THE RESPECTIVE COEFFICIENT OF CORRELATION BETWEEN THE ABSENCE OF REVERSAL TENDENCY AND EACH OF THE OTHER FACTORS*

Subjects	Index of Reversals	Direction of Inspection	Handedness	Per Cent of Right Handedness	Eyedness	Pure and Mixed Eye-Hand Dominance	Space Perception
(1)	97	L-R	L	24	R	M	77
(2)	95	L-R	R	87	R	P	84
(3)	88	L-R	R	99	R	P	66
(4)	86	L-R	R	87	L	M	40
(5)	84	L-R	R	92	R	P	77
(6)	79	L-R	R	88	R	P	68
(7)	77	L-R	R	78	L	M	71
(8)	76	L-R	R	90	R	P	45
(9)	73	L-R	R	97	R	P	54
(10)	72	L-R	R	95	L	M	67
(11)	69	L-R	R	94	R	P	59
(12)	64	L-R	R	90	R	P	89
(13)	64	L-R	R	82	L	M	69
(14)	52	L-R	R	97	R	P	48
(15)	46	L-R	R	88	R	P	56
(16)	41	L-R	R	97	R	P	27
(17)	39	L-R	R	98	R	P	36
(18)	35	L-R	L	8	L	P	47
(19)	33	L-R	R	87	R	P	60
(20)	29	L-R	L	44	L	P	29
	$\sigma = 21$	$r = 0$	$r_{bis} = .35$ (R)	$r = .16$	$r_{bis} = .18$ (R)	$r_{bis} = .53$ (M)	$r = .62$

* Symbols in parentheses below r_{bis} indicate categories with the higher means.

Table I shows the index of reversals, handedness, per cent of right-handedness, eyedness, eye-hand dominance, and space perception ability of each subject, with the respective coefficient of correlation between each factor and the absence of reversal tendency. It will be seen that the only coefficients of correlation that are of sufficient magnitude to be statistically significant at the five per cent level of confidence are those for eye-hand dominance and space perception. The significant correlation for eye-hand dominance is for mixed dominance and absence of reversal tendency.

Table II shows the index of reversal tendency and the results for each factor tested by the Keystone Ophthalmic Telebinocular and Betts' Visual Sensation and Perception Tests for each subject, with the respective biserial coefficients of correlation between each of these factors and the absence of tendency to reversals. It will be seen that none of these coefficients of correlation is of sufficient magnitude to be significant at the five per cent level of confidence.

CONCLUSIONS

The implications of the above findings for the various theories of the cause of reversals in reading considered in this study will be taken up in the order in which these theories are presented in the introduction. These findings will be used to support or negate the various theories, on the assumption that there should necessarily be a significant direct relationship or association between a proposed causal factor and the phenomenon in question. For example, a non-significant coefficient of correlation or a significant negative coefficient between the tendency to reversals and a specific factor which has been proposed as a cause of reversals would strongly suggest that this is not a cause. Likewise, a significant positive coefficient of correlation between this tendency and such a factor would support the theory that it is a cause.

It will be recalled that Cole (*4*) attributed reversals to immaturity. In doing so she was referring to a visual difficulty which she found to be outgrown by the age of seven or eight. She says, "All children reverse letters until they are five or six years old, and some do not outgrow the immature vision which causes these reversals until they are seven or eight." None of the subjects

TABLE II.—INDEXES OF REVERSAL TENDENCY AND RESULTS FOR EACH FACTOR TESTED BY BETTS' VISUAL SENSATION AND PERCEPTION TESTS FOR EACH SUBJECT, WITH THE RESPECTIVE COEFFICIENTS OF CORRELATION BETWEEN EACH OF THESE FACTORS AND THE ABSENCE OF TENDENCY TO REVERSALS*

Subjects	Index of Reversals	Total Test	Two Eyed Vision	Distance Fusion	Two Eyed Efficiency	Left-Eye Efficiency	Right-Eye Efficiency	Vertical Balance	Coördination Level	Distance Lateral Balance	Reading Lateral Balance	Reading Fusion	Reading Sharpness of Image	Distance Sharpness of Image
(1)	97	−	+	+	+	+	+	+	−	+	+	+	+	+
(2)	95	−	+	+	+	+	+	+	+	+	+	+	−	+
(3)	88	−	+	−	+	+	+	+	−	−	+	+	+	−
(4)	86	−	+	−	+	+	+	+	+	+	+	−	+	−
(5)	84	−	+	+	+	+	+	+	+	+	+	+	+	−
(6)	79	−	+	+	+	+	+	+	−	+	+	−	+	−
(7)	77	−	+	−	+	+	−	+	−	−	−	−	+	+
(8)	76	−	+	+	+	+	+	+	+	+	+	−	+	−
(9)	73	+	+	+	+	+	+	+	+	+	+	+	+	+
(10)	72	−	+	−	+	+	+	+	+	+	+	−	+	−
(11)	69	−	+	+	+	+	+	+	+	+	+	+	+	−
(12)	64	−	+	−	+	+	+	+	+	−	−	+	+	+
(13)	64	−	+	+	+	+	−	+	+	−	−	+	+	−
(14)	52	−	+	−	+	+	+	+	+	+	+	+	+	−
(15)	46	−	+	−	+	+	+	+	+	+	+	−	−	+
(16)	41	−	+	+	+	+	+	+	+	−	+	−	−	−
(17)	39	−	+	−	+	+	+	+	+	+	+	+	+	−
(18)	35	−	+	+	+	+	−	+	−	+	+	+	+	−
(19)	33	+	+	+	+	+	+	+	+	+	+	+	+	+
(20)	29	−	+	+	+	+	+	+	+	+	+	+	−	−
	$\sigma = 21$	$r_{bis} = .32$ (−)	$r_{bis} = .00$	$r_{bis} = .03$ (−)	$r_{bis} = .00$	$r_{bis} = .00$	$r_{bis} = .19$ (+)	$r_{bis} = .00$	$r_{bis} = .38$ (−)	$r_{bis} = .07$ (−)	$r_{bis} = .10$ (−)	$r_{bis} = .14$ (−)	$r_{bis} = .41$ (+)	$r_{bis} = .20$ (+)

* Symbols in parentheses below r_{bis} indicate categories with the higher means.

used here could be included in this category nor could any of them be considered to be visually immature, as the youngest was twenty-nine years of age, yet all of them committed reversals in reading the experimental symbols in three to seventy-one per cent of the test words attempted. Betts (*1*) does not identify any particular aspect of immaturity which he associates with reversals but he believes that the period ends at about seven and one-half years and that any tendency to commit reversals can be overcome by this time by training in word perception. But here we have seen an unselected group of superior adult readers commit numerous reversals.

A zero coefficient of correlation was obtained between the results on the Direction of Inspection Test and the Indexes of Reversals, suggesting that there is no relationship whatever between the tendency to commit reversals in reading and the tendency to transfer to reading the techniques or habits of studying or recognizing other objects (Gates (*8*)). All subjects viewed the objects in the Direction of Inspection Test from left to right without exception, yet all committed reversals.

Although all of the coefficients of correlation secured here were computed on the absence of reversal tendencies and the absence of particular factors for statistical convenience, they will be reported here in terms of their reciprocals—the presence of reversal tendencies and the presence of particular factors—in order to make discussion of them more understandable.

A non-significant coefficient of correlation was obtained between tendency to reversals and left-handedness (.35) and per cent of left-handedness (.16). These findings agree with the findings of Woody and Phillips (*25*) and Hildreth (*13*), that left-handedness is not a causative factor in reversals in reading.

A correlation coefficient that is significant at the five per cent level of confidence (.53) was obtained between tendency to reversals and pure cerebral dominance, suggesting that the combination of right-handedness and right-eyedness or left-handedness and left-eyedness, rather than mixed eye-hand dominance, is a causative factor in reversals. In negating the theory of mixed cerebral dominance being a causative factor this finding agrees with the findings of Hildreth (*12*), Gates and Bond (*11*), and Monroe (*18*), and in finding in addition a higher

incidence of reversals among pure dominants it agrees with the findings of Witty and Kopel (*23*).

A highly significant coefficient of correlation (.62), significant at the one per cent level of confidence, was secured between the results on the Survey of Space Relations Ability and the indexes of reversal tendency, thus supporting the theory that reversals in reading are due to difficulties in space relations, confusion of figure-and-ground relationships (Krise (*15*)).

As stated above, failure to establish hemispheral dominance (Orton (*19*)) and verticalization (Fabian (*6*)) were not investigated in this study.

Non-significant coefficients of correlation were secured between the indexes of reversals and the results from all of the tests included in the Betts' Visual Sensation and Perception Tests (Total Test .32, Two-Eyed Vision .00, Distance Fusion .03, Two-Eyed Efficiency .00, Left-Eyed Efficiency .00, Right-Eyed Efficiency .19, Vertical Balance .00, Coördination Level .38, Distance Lateral Balance .07, Reading Lateral Balance .10, Reading Fusion .14, Reading Sharpness of Image .41, Distance Sharpness of Image .20), suggesting that visual deficiency, as measured by these tests, is not a causative factor in reversals in reading. Two subjects passed all of these tests, yet committed reversals in reading. All subjects passed most of the separate tests, yet all committed reversals in reading. The 'various visual defects' which Gates (*9*) feels may predispose the pupil toward reversals—vague images and eye-muscle imbalance—were measured by the separate tests Reading Sharpness of Image, Distance Sharpness of Image, Vertical Balance, Distance Lateral Balance, and Reading Lateral Balance with the above-mentioned non-significant results, and no case of alternating vision—alternating eye-dominance—was found in the group studied.

Theoretical Considerations.—The fact that none of the subjects used in this study committed reversals while reading the familiar reversible symbols in the initial test of tendency to reversals, and all subjects committed reversals in reading the relatively unfamiliar reversible experimental symbols in from three to seventy-one per cent of the test words attempted, suggests that the tendency to reversals is present in varying degrees in all people at all ages, and that the so-called 'overcoming' or 'growing

out of' this tendency is only learning to prevent its operation in dealing with specific materials. If acceptable, this conclusion has implications for remedial work with reversals in reading. It suggests that when a pupil is found to be committing reversals in his reading his remedial instruction should be confined to the specific letters and words he tends to reverse, rather than to secondary materials such as digits and nonsense symbols, or to attempts to overcome a general deficiency or establish a particular capacity, such as particular visual habits, motor coördination, and auditory discrimination. The highly significant coefficient of correlation secured between tendency to reversals and lack of space perception ability, and the fact that most children eventually learn to prevent the tendency to reversals from operating in their reading, suggest that the problem in reversals is one of lack of familiarity with the figure-ground relationship presented by the letters and words reversed. This, in turn, suggests that remedial work in reversals needs to be centered on the up-and-down, backwards-and-forwards, and top-and-bottom characteristics of letters and words reversed. For example, *p*, *b*, *q*, and *d* all have a backbone and a loop; in *p* the loop is forward at the top, in *b* it is forward at the bottom, in *q* it is backward at the top, and in *d* it is backward at the bottom.

The writer is at a loss to find any implication, especially in remedial work, in the significant coefficient of correlation secured between pure eye-hand dominance and tendency to reversals. Pure eye-hand dominance appears not to be related to lack of space perception ability, the only other factor found to be significant, as the coefficient of correlation between these two is not significant (.29). It appears that more research would need to be done in this particular area before any conclusions could be drawn.

BIBLIOGRAPHY

1) Emmet Albert Betts. *The Prevention and Correction of Reading Difficulties*, Evanston, Illinois: Row, Peterson and Company, 1936, pp. 300, 334–335.

2) H. W. Case and Floyd Ruch. *Survey of Space Relations Ability*, Los Angeles: California Test Bureau, 1944.

3) ———. *Manual, Survey of Space Relations Ability*, Los Angeles: California Test Bureau, 1944.

4) Luella Cole. *The Improvement of Reading*, New York: Farrar and Rinehart, Inc., 1938, pp. 37, 284.

5) Allen L. Edwards. *Statistical Analysis for Students in Psychology and Education*, New York: Rinehart and Company, Inc., 1946, pp. 185–189.

6) A. A. Fabian. "Vertical Rotation in the Visual-Motor Performance—Its Relation to Reading Reversals," *J. educ. Psychol.*, 1945, 36, 129–154.

7) Grace M. Fernald. *Remedial Techniques in Basic School Subjects*, New York: McGraw-Hill Book Company, Inc., 1943, p. 177.

8) Arthur I. Gates. *The Improvement of Reading*, New York: Macmillan Company, 1935, pp. 338, 313.

9) ———. *The Improvement of Reading*, New York: Macmillan Company, 1947, p. 313.

10) ——— and C. C. Bennett. *Reversal Tendencies in Reading*, New York: Bureau of Publications, Teachers College, Columbia University, 1933, pp. 20, 4, 6.

11) ——— and Guy L. Bond. "Relation of Handedness, Eye-sighting, and Acuity Dominance to Reading," *J. educ. Psychol.*, 1936, 27, 450–456.

12) G. H. Hildreth. "A School Survey of Eye-hand Dominance," *J. appl. Psycho.*, 1945, 29, 83–88.

13) Gertrude Hildreth. "The Success of Young Children in Number and Letter Construction," *Child Development*, 1932, 3, 1–14.

14) Ralph W. House. *The Diagnosis, Remedy, and Prevention of Reading Difficulties.* Practical Problems in Education, No. 3, New York: Silver Burdett Company, 1938, p. 15.

15) E. Morley Krise. "Reversals in Reading: A Problem in Space Perception?" *The Elementary School Journal*, 1949, 49, 278–284.

16) C. E. Lauterbach. "The Measurement of Handedness," *Pedagogical Seminary*, 1933, 43, 207–212.

17) C. M. McCullough, R. Strang and A. E. Traxler. *Problems in the Improvement of Reading*, New York: McGraw-Hill Book Company, Inc., 1946, 253.

18) Marion Monroe. *Children Who Cannot Read*, Chicago: University of Chicago Press, 1931, pp. 84–86.

19) S. T. Orton. "Specific Reading Disability—Strephosymbolia," *J. Amer. Medical Asso.*, 1928, April 7, pp. 1095–1099.

20) M. J. Van Wagenen and August Dvorak. *Diagnostic Examination of Silent Reading Abilities*, Minneapolis: Educational Test Bureau, Inc., 1939.

21) ———. *Manual, Diagnostic Examination of Silent Reading Abilities*, Minneapolis: Educational Test Bureau, Inc., 1939.

22) David Wechsler and Myrtle L. Pignatelli. "Reversal Errors in Reading: Phenomena of Axial Rotation," *J. educ. Psycho.*, 1937, 28, 215–221.

23) Paul A. Witty and David Kopel. "Sinistral and Mixed Manual-ocular Behavior in Reading Disability," *J. educ. Psycho.*, 1936, 27, 119–134.

24) L. S. Wolfe. "Experimental Study of Reversals in Reading," *American J. Psycho.*, 1939, 52, 533–561.

25) Clifford R. Woody and Albert J. Phillips. "The Effects of Handedness on Reversals in Reading," *J. educ. Research*, 1934, 27, 651–662.

A STATISTICAL REPORT ON A REVISION OF THE MIAMI INSTRUCTOR RATING SHEET

A. W. BENDIG

University of Pittsburgh

In a previous paper (*1*) the author reported on the use of the Miami University Instructor Rating Sheet originally constructed by Crannell (*3*) in the evaluation of instructors in introductory psychology. While the fourteen scales significantly reflected the individual differences in the teaching characteristics of ten instructors, a marked skewing of the total ratings on most of the scales was noted (*1*). Ratings tended to bunch at the favorable end of the scales and an 'end effect' was noted in which the most unfavorable category of the five choices for each scale tended not to be used by the students. It may be argued that this skewing toward the favorable end of the scales reflects the general teaching excellence of the ten rated instructors. However, the similar skewing evident in a table presented by Crannell (*3*, p. 9) where the ratings were of instructors from many different departments and of instructors of differing levels of teaching sophistication suggests that the skewing is a function of the construction of the scales. From a statistical viewpoint the piling up of ratings at one end of the scales would tend to (a) reduce the ability of the scales to discriminate the individual differences in teaching ability between the instructors rated, and (b) make the application to the ratings of statistics based upon assumptions of normality of distribution (t-test, analysis of variance, some correlational techniques) highly debatable. The present study is concerned with a revision of the Miami scales designed to reduce the skewing of the ratings of introductory psychology instructors and to improve the discrimination of the scales.

THE REVISION

In revising the scales with marked skew, four major changes were made in the stems and choices of the items: (a) Scales in which the mean ratings were near the favorable end of the scale and the most unfavorable category was ignored were modified by eliminating the most unfavorable category and adding at the

favorable end a category whose wording appeared to describe the instructor more favorably than a previous favorable category. This was done to all scales with the exception of scales 6, 11, and 14 (see *3*, p. 8 for original wording). (b) Scales in which the wording of the teaching characteristic to be rated seemed to be 'double-barrelled,' i.e., call for rating on two different teaching behaviors that are not on the same continuum, were revised to achieve a more uni-dimensional scale. Scales 11 and 12 were changed in wording. (c) When the favorable ends of rating scales are all at the same side of the page it encourages a constant motor tendency to check ratings at one side of the page and discourages a careful reading of each scale (*7*, p. 271). Therefore, seven of the fourteen scales were reversed in the order of the choices on the scales. The seven scales to be reversed were selected by drawing numbers from a table of random numbers (*5*, p. 340). The scales selected for reversal in this fashion were numbers 1, 2, 4, 6, 9, 11, and 12. (d) The form of the scales was changed from a graphic scale with the choice-points spread horizontally across the page to the more usual multiple choice type of item with a question stem and five choices arranged vertically under the stem. This step was taken because the students to do the rating were more familiar with this type of item from their contacts with the objective achievement tests in the course. All scales were changed in this manner.

The revised scales[1] were mimeographed and standard IBM machine-scored answer sheets provided for the students' responses.

PROCEDURE

As part of a larger research program the revised Miami scales were given to all students in daytime sections of introductory psychology at the University of Pittsburgh near the end of the Spring semester 1950–51. The sample consisted of two hundred forty students in eight sections taught by six instructors. Details as to the course and educational characteristics of students in this course can be found in a previous paper (*1*). Due to the experimental requirements of the research noted above the scales could not be given to the students anonymously. The students were

[1] Available from the author upon request.

requested to sign their names to the ratings. However, this lack of anonymity was at least partially offset by having the ratings administered and collected by two graduate students conducting the research and not by the introductory psychology instructors themselves as in previous studies. These researchers[1] carefully explained to the students that they were being asked to sign their names as part of a research project, that their ratings would in no wise affect their grade in the course, and that their instructor would never see their individual ratings. It is, of course, impossible to determine in this instance whether the lack of anonymity affected the ratings, but the clinical evidence of students' comments and their lack of refusal to sign suggest that lack of anonymity had little if any effect on the ratings.

Since the six instructors rated in the Spring semester on the revised Miami scales had also been rated on the original Miami scales the previous semester a comparative study of the statistical characteristics of the two sets of scales was possible. The six instructors rated were six of the ten instructors reported upon in a previous paper (*1*). They can be identified as instructors A, B, D, F, I, and J analyzed in another study (*2*). The other four instructors of the previous semester were either teaching other courses than introductory psychology during the Spring semester or were employed in research and not teaching.

The ratings given the six instructors on the original (O) Miami scales during the Fall, 1950–51, semester were compared with the ratings given the same six instructors on the revised (R) scales in the Spring, 1950–51, semester. Four statistical aspects of each scale were compared: (a) the means of the scales, (b) the standard deviations, (c) the skewness of the scale ratings, and (d) the discrimination of the scales. Since the scales yielded skewed data the differences between the means of the O and R versions of the scales could not be evaluated for significance by the conventional t-test. Instead, Festinger's test for the means of samples from positively skewed populations (*6*) was used. The standard deviations of the O and R editions were compared by squaring each and determining the ratio of the resulting variances. These variance ratios were evaluated as to significance by con-

---

[1] My thanks are extended to Harold Russell and Richard Krumm for collecting these data.

TABLE 1.—COMPARATIVE STATISTICS ON ORIGINAL (O) AND REVISED (R) SCALES ON THE MIAMI UNIVERSITY INSTRUCTOR RATING SHEET

Scale	Number of Cases (N)		Mean (M)			Standard Deviation (σ)			Skewness (g_1)			Epsilon (ϵ)	
	O	R	O	R	(F)	O	R	(F)	O	R	(D)	O	R
1[1]	305	240	2.51	3.19	1.45**	0.94	1.09	1.33*	0.63**	0.12	0.51*	0.30	0.60
2[1]	320	240	2.48	3.53	1.71**	1.19	1.08	1.21	0.20	−0.64**	−0.84**	0.54	0.60
3[1]	321	240	2.30	2.46	1.12	0.70	0.85	1.49**	0.33**	−0.73**	0.43*	0.20	0.16
4[1]	324	240	1.58	2.60	2.76**	0.62	0.83	1.82**	1.46**	0.21	1.25**	0.83	0.35
5[1]	319	240	2.14	2.83	1.61**	0.83	1.08	1.70**	0.97**	0.70**	0.27	0.30	0.27
6	322	240	1.85	2.45	1.71**	0.98	0.86	1.31*	0.88**	0.93**	−0.05	0.38	0.54
7[1]	320	240	1.83	2.22	1.47**	1.09	1.19	1.19	1.11**	0.61**	0.50*	0.21	0.25
8[1]	320	240	2.58	2.84	1.16	0.96	1.13	1.38*	0.64**	0.51**	0.13	0.21	0.23
9[1]	323	239	2.14	3.04	1.79**	1.26	1.34	1.12	0.75**	−0.08	0.83**	0.11	0.20
10[1]	322	240	1.85	2.37	1.61**	0.87	1.14	1.75**	0.59**	0.77**	−0.18	0.33	0.46
11[1]	320	240	2.23	2.51	1.23	0.96	1.17	1.50**	0.54**	0.43**	0.11	0.36	0.34
12[1]	320	240	1.97	2.60	1.65**	0.69	0.88	1.64**	0.54**	0.03	0.51*	0.36	0.48
13[1]	321	240	2.40	2.88	1.34*	1.07	1.24	1.35*	0.35*	0.05	0.30	0.25	0.21
14	315	237	2.59	3.16	1.36*	0.84	1.02	1.49**	−0.02	0.10	−0.12	0.17	0.43
Fiducial Limits .05					1.28			1.30	0.269	0.310	0.410		
Fiducial Limits .01					1.41			1.44	0.353	0.408	0.539		

[1] Scale changed in wording on revision

* Significant at the .05 level

** Significant at the .01 level

sulting a two-tailed F table (*4*, pp. 311 and 313) with the caution that since the basic data are not normally distributed the fiducial limits read from the F table probably are not too precise. The measure of skewness computed was Fisher's g_1 (*8*, p. 27). These measures of skewness for both O and R scales were separately tested against the null hypothesis of zero skewness by using the standard error formula given by McNemar (*8*, p. 61) and the significance of the difference between the skewness of O and R scales determined by computing standard errors of the difference between this measure of skewness. The unbiased curvilinear correlation coefficient, epsilon, was computed for each scale as detailed in a previous paper (*1*).

RESULTS

The summarized results can be found in Table 1. Eleven of the fourteen means are significantly larger on the R scales and the remaining three scales show insignificant differences between the O and R versions. Eleven of the standard deviations are significantly greater among the R scales, three are not significantly different from their O counterparts, and one is significantly smaller. As to skewness, five of the scales are significantly less skewed among the R scales, seven are not significantly different in the amount of skewness, and two of the revised scales are significantly more skewed than the original Miami scales. Nine of the R scales show larger epsilon values and five show smaller correlation magnitudes.

DISCUSSION

In discussing the statistical comparison of the original Miami scales and the revised edition, two assumptions must be made. One is that the six instructors rated by their students on the original Miami scales in the Fall semester of 1950–51 and on the revised Miami scales in the Spring semester of 1950–51 did not significantly change in the behaviors measured by the fourteen scales from one semester to another. The second assumption is that the composition of students in introductory psychology classes did not vary significantly from semester to semester.

In connection with the first assumption it should be noted that if any significant change took place in the teaching effectiveness

of these six instructors it must have been in the direction of lessened effectiveness, since all of the scales on the revised edition had means higher, or more toward the unfavorable end of the scales, than did the original scales. Any increase in teaching effectiveness in the Spring 1950–51 semester would have had the effect of reducing the demonstrated differences between the original and revised scales. As to the second assumption, enrollment information indicates that the only difference between Fall and Spring semesters is in a slightly smaller percentage of freshmen and a correspondingly greater percentage of senior undergraduates enrolled in introductory psychology.

From the results in Table 1 the general conclusion can be drawn that the majority of the changes made in the original Miami scales resulted in more desirable statistical characteristics. This is not to deny that some of the changes were not successful in achieving the stated aims, for example, scales 2 and 3 were significantly more skewed in the revised edition and scale 6 had a significantly smaller variability as a result of the changed conditions. However, scales 1, 4, 7, 9, and 12 appear to be more useful in the revised edition than in the original version. Scales 3, 5, 6, 7, 8, 10, and 11 will have to be further revised on subsequent usages of the Miami scales in evaluating introductory psychology instructors. In these future editions it is planned to use Crannell's original wording for scales 2 and 14, the present revised wording for scales 1, 4, 9, 12, and 13, and to again revise the remaining scales in the light of the original and revised wording, both of which versions proved unsatisfactory.

SUMMARY

The fourteen scales of the Miami University Instructor Rating Sheet were revised in wording in an attempt to reduce the positive skewing of ratings. Six instructors in introductory psychology were rated by two hundred forty undergraduate students and the revised scales compared with the original Miami scales used the previous semesters with the same instructors. The scales were compared in terms of means, standard deviations, skewness, and discriminatory ability. The original wording of two scales and the revised wording of five scales appeared to give satisfactory distributions of ratings. The remaining seven scales still yield significantly skewed ratings and will again have to be revised.

REFERENCES

1) A. W. Bendig, "The Use of Student Rating Scales in the Evaluation of Instructors in Introductory Psychology." *Journal of Educational Psychology.* XLIII, (1952) 167–175.

2) ———, "An Inverted Factor Analysis Study of Student-Rated Introductory Psychology Instructors." *Journal of Experimental Education.* (in press)

3) C. W. Crannell, "An Experiment in the Rating of Instructors by their Students." *College and University,* XXIII (1948) 5–11.

4) W. J. Dixon, and F. J. Massey, Jr., *Introduction to Statistical Analysis.* New York: McGraw-Hill, 1951.

5) A. L. Edwards, *Statistical Analysis.* New York: Rinehart, 1946.

6) L. Festinger, "A Statistical Test for Means of Samples from Skew Populations." *Psychometrika,* VIII (1943) 205–210.

7) J. P. Guilford, *Psychometric Methods.* New York: McGraw-Hill, 1936.

8) Q. McNemar, *Psychological Statistics.* New York: Wiley, 1949.

SOME PERCEPTUAL FACTORS MEASURED BY MOTION PICTURE TESTS

BENJAMIN FRUCHTER and WILLIAM W. MAHAN

The University of Texas

Motion pictures have a number of advantages for group testing not possessed by paper-and-pencil tests. Some of these advantages, which have been pointed out by Gibson and his associates,[1] are as follows:

1) *Movement.*—In contrast to a printed test a motion picture test can have the characteristic of movement or animation. Two types of movement can be represented. First, there may be movement of one or more objects when judgments of velocity or direction or motion are required. The second type is movement of the observer himself. This is obtained by having the objects remain stationary and the camera in motion during the filming of a sequence. This latter type of motion is especially analogous to flying situations in which (e.g., in landing) the observer moves relative to the ground.

2) *Temporal Sequence.*—This characteristic is especially useful for tests of memory, perception, etc., where the order of presentation, time of exposure and control of the sequence of the examinee's behavior is important. In other words the motion picture projector is used as an exposure device for presenting the material to a group simultaneously.

3) *Pacing.*—Since each item can be timed separately the tempo at which the work is done can be controlled, and is a known quantity. Also, the length of exposure of each successive item can be shortened as the test progresses, if it is desired to measure speed directly.

4) *Realism.*—Compared with photographs or verbal descriptions, a much greater illusion of realism can be created by means of motion pictures. Realistic views of scenes involving movement or action can be presented.

These advantages are especially important for tests of perception since much greater control of the perceptual process is possi-

---

[1] J. J. Gibson, Editor, "Motion Picture Testing and Research," *AAF Aviation Psychology Program Research Reports*, No. 7, Washington, D. C.: U.S. Government Printing Office, 1947, pp. 267.

ble when the material is presented on the screen instead of in a test booklet.

To balance these advantages there are several disadvantages to the use of motion picture tests. For one thing the production of films requires both professional skills and expenses. Also special equipment such as projectors, screens and window blinds are required. Consequently, this method of presenting tests should be applied only for testing functions which cannot be adequately measured by printed tests.

DESCRIPTION OF TESTS AND POPULATION

During World War II a psychological test film unit was set up to develop motion-picture tests of perceptual skills considered critical for flying. In taking off, maneuvering, and landing an airplane the pilot must make judgments from perceptual cues coming from outside of the airplane and from dials in the cockpit. Likewise the bombardier in sighting a target has to use perceptual cues in order to line up his target, often from a height and distance of several miles. Although the tasks mentioned are relative to military flying it should be obvious that similar perceptual cues are used in many other military and civilian situations.

This study will concern itself with nine experimental, perceptual motion picture tests which were developed by the aforementioned motion-picture test unit to measure certain perceptual skills.

Two of the tests were constructed as measures of the ability to put together successive perceptions, i.e., the ability to recognize a total pattern after having seen only parts of it at a time. In one of the tests, *Successive Perception—1*, there are forty-five items in each of which a slot moves down from the top to the bottom of the screen, successively exposing parts of a design. The examinee's task is to identify the complete design from among five choices. The designs become more complex as the test progresses.[2]

The second test in this area, *Successive Perception—2*, consists of fifty items in each of which a moving black spot traces an imaginary path during a ten-second interval. The examinee's task is to recognize the path from among five choices.

Two tests of the ability to perceive slight movement were

[2] For a more detailed description of these tests including sample items, time limits, distribution and other statistics see Ibid., Chap. 5.

included. They present tasks somewhat analogous to what a bombardier does in sighting through a bomb sight. In the *Minimal Movements* test a black spot appears in one of the four quadrants of a circle. It barely moves. The examinee's task is to judge the extent and direction of movement. The spot appears in different quadrants for different items and the items vary with respect to the speed of motion and the length of time of exposure of the spot (two–three seconds).

In the *Drift Direction* test a circle with perpendicular hairlines is shown. A black spot appears in one of the four quadrants of the circle. The spot is stationary for the initial two seconds. The spot then moves slowly but noticeably, either up, down, right or left. The examinee's task is to determine whether, in addition to the main direction of movement, the spot is also drifting toward or away from a hairline. There are sixty items.

Two tests of multiple perception were produced. In the *Integration of Attention* test, five schematic instrument dials are projected on the screen at the same time. The examinee must observe all dials continuously and record simultaneous combinations of wrong dials. The *Flexibility of Attention* test presents the same five dials as in the *Integration of Attention* test. The examinee's task is to distribute his attention over these dials and indicate which have pointers in the 'danger' zone.

In addition to the tests described above three other motion picture tests, each of a different perceptual function, were constructed. *Plane Formation* is a test of quickness of perception of a visual pattern and the ability to reproduce it accurately. The task of the examinee is to reproduce the pattern made by five planes by marking the corresponding sections of a grid on an answer sheet.

In the *Landing Judgment* test the examinee is required to judge at which of five spots on an airport runway his guide is aimed and to record his choice on an answer sheet. The view shown in the film is how the runway would look from an airplane gliding in towards it. There are sixty items in the test.

The *Estimation of Relative Velocities* test requires a judgment of the relation between two velocities. Two planes travel in parallel paths at different speeds. The examinee's task is to estimate at which of five indicated points the faster of the two planes would overtake the slower. There are fifty trials.

The tests were administered to 2,281 unclassified aviation students, along with a large battery of paper-and-pencil tests of perceptual and other abilities. The group had been screened on the Air Force Qualifying Examination and had the equivalent of a high school education. Almost all were either eighteen or nineteen years of age and had passed the visual acuity requirements for aviation students (20-20 vision).

THE ANALYSIS

The distributions of test scores were normalized and intercorrelated by the product-moment method. The correlations were factored to three centroid factors. Table I gives the correlations, centroid loadings, communalities, and also the orthogonal rotated loadings of the three factors.

TABLE I.—INTERCORRELATIONS, CENTROID AND ROTATED FACTOR LOADING FOR NINE MOTION PICTURE APTITUDE TESTS

	Correlations								Centroid Factor Loadings			Rotated Factor Loadings			
	1	2	3	4	5	6	7	8	I	II	III	I	II	III	h^2
1. Successive Perception I									54	−23	−08	*55*	20	09	35
2. Minimal Movements	20[a]								49	36	−23	11	*59*	26	43
3. Successive Perception II	42	22							61	−38	−08	*70*	16	04	52
4. Landing Judgment	11	23	21						33	10	−24	17	*38*	05	18
5. Integration of Attention	19	31	24	07					47	18	23	22	16	*48*	30
6. Plane Formation	31	12	48	14	20				57	−32	09	*63*	04	18	43
7. Flexibility of Attention	26	25	28	11	37	27			54	14	26	29	13	*52*	37
8. Drift Direction	23	32	16	21	15	26	21		45	12	−19	24	*41*	14	25
9. Estimation of Relative Velocities	13	09	07	10	13	15	03		23	05	15	13	03	*24*	08

[a] All entries to two decimal places. Decimal points omitted.

Factor I is defined by high loadings for the following tests:

Test	Loading
3. Successive Perception II	.70
6. Plane Formation	.63
1. Successive Perception I	.55

The function which seems to be common to these three tests is the ability to perceive and recognize a pattern and consequently the factor is tentatively identified as 'pattern perception.' Roff[3] has also analyzed the nine motion-picture tests discussed here as part of a larger battery of seventy tests. He obtained a similar factor which he identified as 'sequential perception.' Since there is no sequential perception involved in the Plane Formation test the present designation may be more appropriate. Roff[4] also has observed that the three tests loaded on this factor were also loaded on a perceptual-speed factor, indicating a possible relationship between the two factors.

Factor II is defined by high loadings on the following tests:

Test	Loading
2. Minimal Movements	.59
8. Drift Direction	.41
4. Landing Judgment	.38

All three tests require judgments as to amount and direction of movement. In the two tests with highest loadings on the factor the amount of motion is at the threshold of perception. The factor is labelled 'movement detection.'

Factor III is defined by the following tests:

Test	Loading
7. Flexibility of Attention	.52
5. Integration of Attention	.48
9. Estimation of Relative Velocities	.24

The two attention tests require the examinee to divide his attention among several dials while making certain observations. The *Estimation of Relative Velocities* test had low correlations with the other motion-picture tests and most of its variance is specific, unaccounted for variance. It was included in the above list since it had its highest loading on this factor. Perhaps, in order to judge the relative velocities of the two airplanes it is necessary to divide one's attention between them somewhat. The common function seems to be 'division-of-attention.'

---

[3] M. F. Roff, *Personnel Selection and Classification Procedures: Perceptual Tests.* USAF School of Aviation Medicine, Randolph Field, Texas, April 1950, pp. 23.

[4] *Ibid*, p. 14.

In the solution by Roff[5] referred to above these three tests appeared on a factor along with a number of pencil-and-paper tests which appeared to require speed of perception in a complex perceptual situation. Roff identified the factor as 'complex perception.'

SUMMARY AND DISCUSSION

Motion pictures have a number of unique advantages over paper-and-pencil tests for group testing. The advantages may be summarized under the headings of movement, sequence, pacing and realism. Since, however, the production and administration of such tests involves considerable skill and expense they should be used only for testing functions which cannot be adequately tested by printed tests. Perceptual abilities are one area where the unique advantages of motion-picture tests are especially useful. Measures of perceptual speed and estimation of velocities can be obtained better by animated pictures whose exposure time can be controlled.

During World War II a test film unit was set up for research on the development and use of motion-picture tests of perceptual skills important in flying training. Nine motion-picture tests of various perceptual functions were produced and administered, along with a battery of experimental paper-and-pencil aptitude tests, to 2,281 aviation students. Centroid analysis of the intercorrelations of the nine motion-picture tests yielded three factors. After rotation the factors were identified as pattern perception, movement detection, and division-of-attention.

---

[5] *Ibid*, p. 13.

THE RELATIONSHIP OF LITERARY APPRECIATION TO CERTAIN MEASURABLE FACTORS

DWIGHT L. BURTON

Florida State University
Tallahassee, Florida

'Literary appreciation' has been a term to conjure with in education. Most teachers of language arts, from the elementary through the graduate school, profess to be teaching, or at least promoting, literary appreciation. Yet there is no particular agreement on what literary appreciation is; there is rather complete disagreement on how it should be taught, if, indeed, it can be taught at all; and few accepted methods have been developed for measuring it. This study sought to clarify the nature of literary appreciation by discovering its relationships to verbal and nonverbal intelligence, silent reading ability, and socio-economic status.

A series of tests was administered in October, 1950, to one hundred ninety twelfth-grade students in three Minnesota high schools. The tests measured intelligence, silent reading ability, and literary appreciation. A rating scale was used to classify the students according to socio-economic status. Coefficients of correlation were computed among the scores on these various measures.

MEASURING INSTRUMENTS

1) *Intelligence.*—The California Test of Mental Maturity, Advanced Series, was used to measure intelligence. The test yields a language and nonlanguage score, and the reliability coefficient of the total score is consistently in the high 90's.

2) *Silent Reading.*—The Iowa Silent Reading Test, Form Am, was the measure of silent reading skill. The reliability coefficient of the total score is variously reported in the low 90's.

3) *Literary Appreciation.*—(a) Carroll Prose Appreciation Test, High School Form—This test, which is described in detail elsewhere (*2*), measures discrimination in prose style, one aspect of literary appreciation. Each of the twelve items of the test require students to rate, in order of quality, four excerpts from

published prose works. The reliability coefficient is .71. (b) Burton Short Story Comparison Test, Form A—This is one of two measures developed by the author. Full discussion of the theory, construction, and validation of this test and the one discussed in (c) below can be found elsewhere (*1*). The Short Story Comparison Test consists of two published short stories followed by a twenty-item multiple choice test which requires a critical comparison of the two stories. One selection represents typical stereotyped, superficial fiction while the other is an example of mature art in the short story. The reliability coefficient of the test, computed by correlating scores on the equivalent forms A and B, is .74. (c) Burton Short Story Choice Test—This is a measure of general maturity of taste in the contemporary short story. Each of the ten test items presents a summary of a short story. The summary stops at a certain point, and three versions are given of how the story could be completed. The student rates these three versions in order of quality. The reliability coefficient, based on the split-halves method, is .83.

4) *Socio-economic Status.*—The Minnesota Occupational Scale was used to assign 'scores' for socio-economic status. The scale provides a seven-category classification of occupations; each student was placed in one of these categories according to the occupation of the head of his household. These 'scores,' which are really categories rather than continuous scores, met the criterion of linearity of regression on scores from the various tests. Thus, ordinary zero order correlation was used to obtain the relationships between the socio-economic level and the various tests of literary appreciation.

RESULTS

The product moment coefficients of correlation among the scores on the various measures are given in Table 1. The correlation coefficients corrected for attenuation are given in parentheses.

INTERPRETATION OF RESULTS

The correlations found between scores on the three appreciation tests and the total IQ score closely parallel those found in other studies (*3*, *4*, *5*). However, the relationship is explained largely by the language factor of the test; there is little or no relationship between nonlanguage intelligence and literary

TABLE 1.—COEFFICIENTS OF CORRELATION OF SCORES ON TESTS OF LITERARY APPRECIATION, SILENT READING ABILITY, INTELLIGENCE, AND SOCIO-ECONOMIC STATUS

	Total IQ	Lang. IQ	Non-Lang. IQ	Iowa Read.	Socio-Econ.	Carroll Test	Choice Test	Comp. Test
Carroll Test	.37	.52	.10	.47	.23	—	.39	.44
	(.45)	(.61)	(.12)	(.58)			(.51)	(.61)
Burton Choice Test	.20	.44	−.03	.27	.32	.39	—	.36
	(.22)	(.50)	(−.03)	(.31)		(.51)		(.46)
Burton Comparison Test	.36	.64	.11	.33	.33	.44	.36	
	(.42)	(.77)	(.13)	(.40)		(.61)	(.46)	—

appreciation of the types measured. This together with the relatively substantial correlations found between the tests of appreciation and the Iowa Silent Reading Test, points to the reading or language factor as the most important in literary appreciation. It is interesting to note that the correlations between intelligence—total, language, and nonlanguage scores—and the Choice Test tend consistently to be somewhat lower than those between intelligence and the other two tests of appreciation. The Carroll and Comparison Tests require much more specific technical judgment about literature than the Choice Test which puts the main premium on recognizing general significance in a story and on rejecting false assumptions about human experience. It seems likely that the narrower and more technical the concept of literary appreciation becomes, the closer it will be related to intelligence. Often, appreciation of literature is thought of as the ability to analyze form and style. This is only one, and perhaps a relatively unimportant, part of literary appreciation. There is some basis probably for believing that important aspects of literary appreciation can be taught even to students on lower levels of intelligence and reading skill.

The results indicate, as other researches have (*4*, *5*), the presence of a separate factor of appreciation; that is, high intelligence and silent reading ability applied to literary materials will not guarantee literary appreciation. Furthermore, the correlations among the three appreciation tests indicate that appreciation of literature tends to be fairly specific. A student's ability to appreciate literature cannot be judged reliably by a single test.

CONCLUSIONS

A study of correlations among scores on tests given to one hundred ninety twelfth-grade students in three Minnesota high schools leads to certain generalizations about the nature of literary appreciation:

1) Verbal intelligence and silent reading ability are important factors in appreciation of literature.

2) Nonverbal intelligence has negligible relationship to the elements of literary appreciation measured by the tests used in this study.

3) There is an important factor of appreciation of literature separate from intelligence and reading skill.

4) Socio-economic background determines to some extent a student's ability to appreciate literature.

5) Literary appreciation tends to be specific rather than general.

REFERENCES

1) D. L. Burton, *A comparison of three methods for teaching appreciation of the short story to twelfth-grade students.* Unpublished doctoral dissertation, Univ. of Minn., 1951.

2) H. A. Carroll, "A method of measuring prose appreciation." *Eng. J.*, 1933, 22, 184–189.

3) ———, "Appreciation of literature and abstract intelligence." *J. ed. Psychol.*, 1934, 25, 54–57.

4) R. K. Speer, *Measurement of appreciation in poetry, prose, and art.* Teachers College, Columbia University, 1929.

5) J. M. Woods, L. Winter, and E. D. Williams, "Tests of literary appreciation." *Brit. J. ed. Psychol.*, 1938, 8, 269–284.

BOOK REVIEWS

S. S. STEVENS, Editor. *Handbook of Experimental Psychology.* New York: John Wiley and Sons, Inc., pp. 1436. $15.00.

This massive handbook supplants Murchison's earlier *Handbook of General Experimental Psychology* published in 1934 and as such serves a genuine need by bringing together under the covers of a single book a great deal of important information of concern to psychologists with the experimental point of view. Since thirty-four psychologists have contributed to the thirty-six chapters in the book there is lack of uniformity, but this is to be expected in an edited work. Of importance is the fact that the authors represent a distinguished group of psychologists, and each contribution is authoritative and scholarly. To be sure, some of the chapters represent broad reviews of the topic assigned while others have presented narrower interpretations representing specific interests and points of view. Some are eclectic and fairly present different points of view while others tend to be polemic and to present evidence that represents the point of view of the author.

This review, addressed to educational psychologists, attempts to point out the sections and chapters which educational psychologists may find of most interest and value. From one point of view, everything of interest to psychologists pertains more or less directly to education; however, the gap between physiology and behavior has not yet been completely bridged and interpretations about the physiological bases for behavior, and even more strongly, deductions of behavior from physiological processes still remain highly speculative. For this reason, the sections on the "Physiological Mechanisms" and on the "Sensory Processes" may have less immediate value to educational psychologists.

I would have expected the section on "Growth and Development" to be of interest to educational psychologists, but four of the chapters deal with neural maturation, prenatal development, physiological development and problems of heredity, and as they have been treated, are of less immediate concern to educational psychologists. In the chapter on "Phylogenetic Comparison" by Nissen there are sections on the motivational aspects of behavior and the cognitive aspects of behavior which should be of interest to those concerned with education.

In the section on "Motivation," by far the most important paper is that by Neal Miller on "Learnable Drives and Rewards." This paper was looked forward to with considerable anticipation because of the promise it held from references to it in the book by Dollard and Miller on *Personality and Psychotherapy*. This reviewer must confess that Miller's chapter was to a degree disappointing partly because he did not go into some of the theoretical questions which were left unsettled in the book referred to. To be sure Miller reviews neatly for us the experimental evidence on secondary or derived drives and rewards, but since Miller has defined drive as any strong stimulus there is confusion between drive, stimulus and goal that leaves the reader somewhat bewildered. Part of this difficulty is merely semantic, but because of the failure to present a defensible theoretical foundation, the clear assessment of the experimental work is lost.

Several chapters in the section on "Learning and Adjustment" are of considerable value to the educational psychologist. Hilgard has prepared a very important chapter outlining methods, procedures and apparatus to be used in the experimental study of learning. Hovland has made a useful review of recent experimental work in "Human Learning and Retention." Perhaps the most significant chapter in the entire volume is that by Spence entitled "Theoretical Interpretations of Learning." In this chapter he presents some of the evidence and arguments for and against various learning theories such as the problem of reinforcement, the two factor theory of learning, the problem of latent learning, of inhibition, of cognition, of mechanism in learning and other theoretical issues. In spite of the fact that Spence has committed himself to a reinforcement position with regard to learning he is fair in presenting evidence for other points of view and for not drawing biased conclusions. He concludes with the following pregnant sentences: "What is needed most at the present time in the study of learning is the continued and persistent pursual of the type of integration of theory and experiment that will lead to the discovery and formulation of laws of learning in terms of environmental and behavioral variables. When such a body of knowledge is available we shall be in a better position to evaluate the adequacy of physiological hypotheses."

Leeper's chapter on "Cognitive Processes" is also ripe with

suggestions for the educational psychologist. Morgan's chapter in the same section dealing with "The Psychophysiology of Learning" is undoubtedly authoritative but one gets the impression that not much more is known than was reported to us several years ago by Lashley concerning the processes in the nervous system during learning. Miller's chapter on "Speech and Language" presents some of Miller's unique contributions with regard to the place that language plays in modern civilization.

There are three chapters in the last small section in the book on "Human Performance" that also concern the educational physiologist. Bechtoldt has written a compact statement of modern principles in differential psychology as they relate to the use of tests in the process of personnel selection. Dael Wolfle has an interesting chapter on "Training" presenting recent experimental work on the influence of the teacher or instructor in arranging learning situations for greater effectiveness. Coming late in the book, there is danger that this chapter may be overlooked. And finally the last chapter in the book by Seashore deals with motor phenomena.

These chapters which have been commented upon are of such significance that the educational psychologist cannot afford to overlook this book, and he should find it a continual source of reference to these pertinent topics of motivation, learning and personnel selection.

PERCIVAL M. SYMONDS
Teachers College, Columbia University

DOROTHY M. ROBERTS. *Leadership of Teen-Age Groups.* New York: Association Press, 1950, pp. 195.

The keynote of this treatise is found in the sentence "Teen-agers are *people*, not *problems*." The book has a twofold purpose: (1) to show that teen-agers are people like ourselves although younger and less experienced; (2) to share with others the ideas on leadership of teen-age groups that have proved acceptable and effective. The four parts of the book deal with: the world of the teen-ager; helping teen-agers organize their groups, leadership of teen-age groups, and program planning with teen-age groups.

The material is a highly condensed, clearly stated, and practical account of principles and procedures which the author has shown through her own experience to be effective. These principles and practices are sound psychologically. All adults who work

with young people in such organizations as churches, schools, clubs and camps will profit from use of the information in this excellent book.

MILES A. TINKER
University of Minnesota

D. H. STOTT. *Delinquency and Human Nature.* Dumferline, Fife, England: Carnegie United Kingdom Trust, 1950, pp. 446.

Delinquency and Human Nature reports an intensive study of a group of delinquents made by a psychologist who was granted facilities for residence in an Approved School in England. Starting in 1945, Mr. Stott interviewed one hundred two children and, in practically every instance, the parents. The contents are recorded in fifteen chapters with emphasis on home background and parental background. Names of typical chapter captions are: "Parental Ill Health," "Parental Quarrels," "Emotional Difficulties Following the Loss of Parent," "Boys Unwanted or Abandoned in Childhood," "Boys Discriminated Against During Childhood," "Unsatisfactory Parents." Each chapter is followed by a summary called 'Argument.' This paper volume also contains three appendices composed of theoretical notes, method in future research and explanations and notes on statistical summary; also, a statistical summary and an index.

The general theory of human nature in this case is definitely non-Freudian—Freudianism being almost practically forbidden. In the first specific chapter on the case studies presented called "Avoidance," we are informed that "the phenomenon of avoidance is, of course, well-known among clinical psychologists. It is the 'repression' of the vulgarized Freudian terminology." Says the author: "I have eschewed the term 'repression' because it is a misleading concept." As might be expected by this reaction to the concept of 'repression' even the interviews when presented do not penetrate into underlying motives very often and are usually presented in clean manner.

The case studies are analyzed in terms of five motives and some supplementary ones. The five main motives considered are: avoidance-excitement, inferiority-compensation, delinquent-attention, resentment directed against parents, removal from home. Classifications are made by the author in terms of introspective reports or from interpretation as it conforms to the

reaction pattern recognized. In some boys, the author tells us, he can detect but a single motive, but in others there is a mixture of as many as four. Where more than one was present no attempt was made to single out the predominant one. The most frequent motive given is avoidance-excitement present in fifty-three cases. Delinquency in these cases took the form primarily of breaking into houses, shops, warehouses, etc. The next most spectacular motive given by the author is represented by forty-two cases in which delinquency was directed as an act of spite, retaliation and resentment against the parents. Delinquent-attention was identified in twenty cases. These are illustrative of the way in which motives are analyzed in this study.

All in all, this paper volume is an interesting illustration of a case study approach to the study of delinquency in relationship to parental and home backgrounds in England which can be useful for comparing cultural differences. Very little attempt is made in this volume to hook up the present study with any other investigations made at other times or in other countries.

H. Meltzer

Psychological Service Center
St. Louis, Missouri

Lecky Prescott. *Self Consistency, A Theory of Personality.* New York: Island Press, 1951, pp. 275.

The present edition of Lecky's *Self Consistency* is more than a revision of the earlier 1945 issue. It includes all the material of the first volume—Lecky's own material as well as Gardner Murphy's preface. But its chief differential characteristic is what Thorne has done to the volume so that in its present form it is really Thorne's presentation and interpretation of Lecky's contribution and its significance as well as a consideration of the nature and motivation of Lecky as a contributor. The content is presented in twelve unevenly organized chapters and two appendices. Thorne has written a biographical sketch of Lecky, a consideration of the nuclear theory of the mind, a critical evaluation of self-consistency and what he refers to in Chapter 13 as "Preventing Failure by Removing Resistance." He also wrote parts of chapters which concern themselves with the topics of validity of personality inventories, the individuality record and clinical approaches to personality study.

The concept of self-consistency as Thorne interprets it is more related to the Adlerian concept of style of living as the determining force rather than any contribution from either Freud or Jung. According to the theory of self-consistency, one seeks experiences which support one's values and avoids, resists or, if necessary, forcibly rejects those which are inconsistent with them. In line with this doctrine of consistency an argument is built up against the doctrine of specificity by May and Hartshorne. Instead, an emerging concept is favored. The inventory method is described in the volume as "both a valid method of gathering the necessary facts and perhaps the only method which can force the recognition of self-consistency and personality." Lecky's concerns with out-growing behaviorism and mechanism of his age, with reacting negatively to the then much talked of character studies of May and Hartshorne, and his concluding that his inventory is the method of choice for studying personality dates the thinking involved as belonging to the late twenties and not in accord with present emphasis as it is developing in the field of clinical psychology. Surprisingly enough, Thorne, who does much interpreting throughout the volume, makes no remarks whatever about Lecky's claim about the inventory being the method of choice for measuring personality organization.

It is interesting that Gardner Murphy who in his original preface to the volume gives Lecky credit as developing in his own way further than Schilder or Kurt Goldstein the conception that the individual must define for himself the nature of that totality which he is, comments on this contribution somewhat more reservedly in his own recent comprehensive volume on *Personality*. In his later *Personality*, Murphy expresses himself as follows: "The urge to consistency can indeed be accepted as a primary mechanism, but if consistency were regularly achieved neurotic tension would be evident only in moments of forced alteration in the design, whereas clinical experience makes it clear that unity is hard to achieve and the struggle to achieve it never subsides."

Lecky's contribution is of interest to many clinicians, and Dr. Thorne deserves our thanks for making it more readily available in a somewhat more intelligible form.

H. MELTZER
Psychological Service Center
St. Louis, Missouri

CHARLOTTE BUHLER, FAITH SMITTER, SYBIL RICHARDSON AND FRANKLYN BRADSHAW. *Childhood Problems and the Teacher*. New York: Henry Holt and Company, 1952, pp. 372. $3.75.

"This book attempts to do three things: to give the teacher an understanding of the dynamics of behavior problems that he may observe, to determine by means of case examples what the teacher may achieve in various circumstances, and to describe the type of problem with which the teacher needs specialized assistance."

It discusses first the teacher and psychology; then the biological and psychological basis and approach to children's problems; third, ways in which teachers can work out the problems not only for children but also work with parents; and, fourth, the contributions of the psychologist to the problems within the school.

The biological and psychological basis emphasizes psychosomatic orientation, the integration of development and adjustment, transformation and change as well as increase and decrease of form and functions, expansion and homeostasis, stability as related to needs and stresses, primary needs, goal directed striving or motivation, mastery, contentment, heredity equipment, total life experiences, maturation and learning, and the desirable outcome in well integrated development and adjustment.

As new concepts in development and clinical psychology, attention is called especially to normality, personality integration, average developmental expectation, self-evaluation, creativity, environmental influences. The educator needs the help of the psychotherapist.

Forty-one cases are used as illustrative material of different kinds of problems and as the means of suggesting desirable psychological and educational procedures. A list of the children with names, ages and page references is given and there are forty-one illustrations, chiefly of drawings made by a large number of the children. An inventory of the cases by problems (pp 358–360) is given under the following headings, main complaint, problem pattern, pseudonym, age, and page references. The index of authors is brief, part of one page (p. 361) and the index of subjects is quite extensive (pp. 363–372).

The book is essentially a series of thoughtful discussions on common and important problems of children and of schools,

without dependence upon any one particular school of psychology, but with the use of contributions from several schools, especially, dynamic and functional, and with some reference to psychoanalytic concepts. The selection of references was limited to a small number that were considered to be helpful but "The selection offered does not deny equal importance to other concepts, problems, methods, and references that could not find place in this presentation." "This book tried to unite developmental and dynamic aspects which usually are held apart."

Many valuable suggestions and much psychological insight may be had from a study of this book.

A. S. EDWARDS
The University of Georgia

C. A. BERTHOLD. *Administrative Concern for Individual Differences,* New York: Bureau of Publications, Teachers College, Columbia University, 1951, pp. 225.

A scale of attitude toward individual differences was constructed and administered to a group of secondary-school administrators. Attitudes and practices as related to individual differences were studied with respect to following areas: instruction, guidance, health, recreation, home background and personality. As such, the book presents a cross section of attitude and current practice. While in general expressions of attitude were found to be more in line with prevailing concepts of good procedure than were the practices in the schools in the sample, the two were found to be positively correlated, particularly in the areas of health, personality, and home background.

J. B. STROUD
State University of Iowa

MILDRED C. TEMPLIN. *The Development of Reasoning in Children with Normal and Defective Hearing.* Minneapolis, Minn.: University of Minnesota Press, 1950, pp. 143.

This study was designed to investigate the influence of restriction of environment upon the development of reasoning in children. Two kinds of restrictions were studied: (1) the effect of loss of hearing, or the intrinsic factor; and (2) residence in institution, the extrinsic factor. The subjects used for this study were eight hundred fifty children from grades five to twelve.

The sample of children selected included: normal day-school hearing subjects, residential hearing subjects, day-school children with defective hearing, and residential subjects with defective hearing. Children with normal hearing in both day and residential schools are compared with children with defective hearing in both day and residential schools. The experiment was designed so that matched groups can be set up for each of the four basic additions. The tests used were the Deutsche questions on physical causality; Long and Welch Test of Causal Reasoning; Brody Non-Verbal Abstract Reasoning Test; and Pintner Non-Language Mental Test. The main conclusion of this relatively well-designed experiment is that restriction of the environment does result in lower scores on the reasoning tests used for the intrinsic factor of hearing loss. But restriction on the extrinsic factor of residence in an institution does not result in such a loss. In the light of the consequences of these experiments the tools and procedures used are evaluated and research suggestions made.

Some more specific problems are also considered in this investigation. Examples are: the maturity of written language of the defective hearing as compared with the hearing subjects; the relation of reasoning to the amount of hearing loss and the age of the onset of hearing loss; and the relation of reasoning to grade placement and intelligence. Dr. Anderson, the director of the Institute for Child Welfare, in his preface justifiably enough rates this particular investigation as one of the most substantial studies in its area. The content of this volume should be of interest not only to research psychologists in the field of hearing but to clinicians and workers in the field of the hard-of-hearing and the deaf.

H. MELTZER

Psychological Service Center
St. Louis, Missouri